Successional and Spatial Patterns of Bacterial Communities in Hydrocarbon-Contaminated Soils and *Populus* rhizosphere

Shinjini Mukherjee

Faculty of Biological and Environmental Sciences
Department of Biosciences
General Microbiology
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in Auditorium 2 at Viikki Infocenter Korona (Viikinkaari 11), on September 12th 2014, at 12 o'clock noon

Helsinki 2014

Supervisor	Docent Kim Yrjälä Department of Biosciences General Microbiology University of Helsinki Finland
Thesis advisory committee	Docent Kirsten S. Jørgensen Marine Research Center Finnish Environment Institute Helsinki, Finland
	Docent Merja Itävaara VTT Technical Research Centre of Finland Espoo, Finland
Reviewers	Professor Jaak Truu Institute of Molecular and Cell Biology University of Tartu Estonia
	Professor Tomas E. Macek Department of Biochemistry and Microbiology Institute of Chemical Technology, Prague Czech Republic
Opponent	Professor George Kowalchuk Institute of Environmental Biology Utrecht University The Netherlands
Custos	Professor Benita Westerlund-Wikström Department of Biosciences General Microbiology University of Helsinki Finland

ISBN 978-951-51-0081-8 (paperback)
ISBN 978-951-51-0082-5 (PDF; http://ethesis.helsinki.fi)
ISSN 1799-7372
Unigrafia
Helsinki 2014
Cover illustration by Dr. Abhishek D. Garg

Table of Contents

Summary

Ever-increasing urbanization and industrialization have led to contamination of a vast numbers of terrestrial sites with petroleum hydrocarbons. Petroleum hydrocarbon pollution has a deleterious impact on biotic and abiotic properties of ecosystem and can thereby affect some valuable ecosystem services. Microbes have the ability to metabolize various components of these harmful contaminants; this unique ability has been harnessed for decades in form of bioremediation and rhizoremediation, with varying success. Advanced knowledge on ecology of microbes in contaminated ecosystems can pave the way for improved design, optimization and monitoring of suitable bio-/rhizoremediation regimes in the clean-up of polluted sites. Recent developments in the molecular microbiology have enabled us studying microbial community structure, function and dynamics at great resolution and precision.

In this thesis, two important dimensions of microbial ecology in polluted ecosystems were explored: temporal and spatial. Succession of microbial communities in contaminated soil and *Populus* rhizosphere were studied in two different experiments. A short-term greenhouse study was conducted to monitor the immediate response of rhizosphere-associated and soil bacterial communities to oil pollution. We further scaled up our study to monitor the bacterial succession during a 2-year field study which also allowed us to analyse the effect of seasonal variation in boreal climate zone. Finally, a case study on an aged creosote-contaminated site located in South-eastern Finland was carried out in order to investigate the spatial patterns of microbial diversity and activity in relation to the heterogeneity of soil chemical parameters. Dynamics and diversity of microbial communities were accessed by employing T-RFLP fingerprinting and 454 pyrosequencing of structural and functional marker genes.

Successional changes in microbial communities could be observed in both our time-series experiments. High resolution sequencing of phylogenetic and catabolic marker genes for microbial community profiling not only enabled us to identify the bacterial groups during different stages of succession but also provided some insights on the structure-function relationship of bacterial communities. A gradual shift from specialist to generalist strategy was observed in the communities of aromatic and aliphatic degraders during the secondary succession in oil pollution. Effect of *Populus* rhizosphere on the general bacterial community structure was masked by the heavy oil pollution but upon careful examination of catabolic gene communities, rhizosphere-prevalent groups were observed. A significant variation in bacterial community structure was observed during the winter months pointing towards a distinct seasonal effect.

Our study on spatial heterogeneity of microbial communities in an aged contaminated site highlighted niche differentiation as the major mechanism regulating bacterial community structure. Geostatistical modelling and spatial prediction brought forward two distinct patterns in geochemical properties - patchy distribution of creosotes and a natural gradient of pH on the polluted site. While most bacterial taxa drastically reduced in abundance in the hotspots of pollution, Proteobacteria clearly dominated these zones. Acidobacteria, on the other hand, responded only to the pH variation irrespective of the differences in pollution levels. Analysing the behaviour of bacterial groups at lower taxonomic levels further clarified the patterns of

niche differentiation created by combined effect of pH and contaminants. The spatial profiles of specific microbial taxa could be used as proxies or indicators for monitoring this polluted site.

The results obtained in this thesis project are not only scientifically interesting but they also find an application in real-time ecological restoration. Both of our studies on bacterial secondary succession were carried out as a part of a phytoremediation project in collaboration with the Finnish Forest Research Institute. Phytoremediation with hybrid aspen is being implemented at the creosote-contaminated site in Luumäki, Finland (2013-). Our results on the spatial heterogeneity of microbial diversity and activity played a great role in the pre-evaluation of this site for remediation. This knowledge on the spatial patterns of microbial diversity will be highly useful in the coming years for the monitoring and evaluation of phytoremediation in the creosote-polluted site.

List of original articles

This thesis is based on the following publications:

I. **Mukherjee S**, Heinonen M, Dequivre M, Sipilä T, Pulkkinen P & Yrjälä K (2013). Secondary succession of bacterial communities and co-occurrence of phylotypes in oil-polluted *Populus* rhizosphere. *Soil Biology and Biochemistry* 58: 188–197.

II. **Mukherjee S**, Sipilä T, Pulkkinen P & Yrjälä K (2014). Secondary successional trajectories of structural and catabolic bacterial communities in oil-polluted soil planted with hybrid poplar. Submitted manuscript.

III. **Mukherjee S**, Juottonen H, Siivonen P, Quesada C-L, Tuomi P, Pulkkinen P & Yrjälä K (2014). Spatial patterns of microbial diversity and activity in an aged creosote-contaminated site. *The ISME Journal* doi: 10.1038/ismej.2014.151.

Authors' contributions

	I	**II**	**III**
Study design	**SM**, MH, TS, PP& KY	**SM**, PP, KY	**SM**, PT, PP, KY
Sampling	MD, MH, **SM**	**SM**	Golder associates**, SM**, PS
Microbial molecular analyses	**SM**, MD	**SM**	**SM**
Other analyses	**SM**, MH	**SM**	**SM**, PS, CLQ
Data handling	**SM**	**SM**, TS	**SM,** HJ
Writing the manuscript	**SM**, KY	**SM,** KY	**SM,** HJ

SM= Shinjini Mukherjee, MH= Mirja Heinonen, MD= Magali Dequivre, TS= Timo Sipilä, KY= Kim Yrjälä, PP= Pertti Pulkkinen, PT= Pirjo Tuomi, PS= Pauli Siivonen, CLQ= Cosme Lloret Quesada, HJ= Heli Juottonen

Abbreviations

2,4-D	2,4-dichlorophenoxyacetic acid
AlkB	alkane monooxygenase
alkB	gene encoding alkane monooxygenase
AlmA	flavin-binding monooxygenase
ARISA	automated ribosomal intergenic spacer analysis
BLAST	Basic local alignment search tool
bp	base pairs
bphC	gene encoding 2,3-dihydroxybiphenyl-1,2-dioxygenase
BTEX	benzene, toluene, ethylbenzene and xylene
cDNA	complementary DNA
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
EDO	extradiol dioxygenase
EST	expressed sequence tag
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
GST	glutathione S-transferase
gyrB	gene encoding the DNA gyrase subunit B
LadA	long-chain alkane hydroxylase
LH-PCR	length heterogeneity polymerase chain reaction
MPN	most probable number
MTBE	methyl terbutyl ether
NADH	reduced nicotinamide adenine dinucleotide
nahC	gene encoding 1,2-dihydroxynaphthalene-dioxygenase
NGS	next generation sequencing
NM-MDS	non-metric multidimensional scaling
NPMANOVA	non-parametric multivariate analysis of variance
OTU	operational taxonomic unit
PAH	poly aromatic hydrocarbon
PCA	principal component analysis
PCBs	polychlorinated biphenyls
PCR	polymerase chain reaction
pMMO	particulate methane monooxygenase
RDA	redundancy analysis
RFLP	restriction fragment length polymorphism
rpoB	gene encoding the beta subunit of RNA polymerase
rpoD	gene encoding the sigma subunit of RNA polymerase
rRNA	ribosomal ribonucleic acid
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sMMO	soluble methane monooxygenase
SSCP	single-strand conformation polymorphism
TCE	trichloroethylene
TGGE	temperature gradient gel electrophoresis
TNT	trinitrotoluene
TPH	total petroleum hydrocarbon
T-RF	terminal restriction fragment
T-RFLP	terminal restriction fragment length polymorphism
VOC	volatile organic compound

1. Introduction

It has long been envisaged that polluted environments can be efficiently remediated through the application of amenable and controllable microbial activities. While this conjecture has been ably supported by some promising laboratory results, the application of pollutant-degrading microbes in field scenarios has mostly ended in disappointment. It has been proposed that for more optimal and rationale-driven remediation, it would be desirable to utilise the inherent catabolic potential of polluted sites. Currently however, there are limitations in harnessing this potential because of the lack of thorough understanding of ecological behaviour of microbial communities and their optimum catabolic capacity in such sites. It is noteworthy that, degradation of contaminants in polluted environments is mainly carried out by a microbial food web rather than a single species. In such situations, the key species and catabolic genes/enzymes involved are often different from those characterised in a laboratory setting. Recent advancements in molecular techniques have enabled us to determine the structural and catabolic microbial diversity at unprecedented resolutions. Harnessing these advancements in molecular microbial ecology would better guide our efforts towards sustainable remediation of polluted sites. In the following chapters, I explore some specific aspects of microbial ecology in hydrocarbon contaminated ecosystems.

1.1. Hydrocarbon contamination in terrestrial ecosystems

Petroleum products are fundamental to our lives in form of transportation fuels, heating and power-generating fuels and many other useful derivatives. The volume of petroleum products used today dwarfs all other chemicals of environmental health and concern. Extraction, refinement and transportation of petroleum have resulted in surface and near-subsurface soil contamination with crude oil, gasoline, diesel and creosote. The composition of released petroleum products varies significantly depending on the source, the physical, chemical and biological changes (collectively referred to as weathering) of the product over time, and differential movement of the components in the environment. Crude oil is a very complex mixture of more than several thousand distinct chemical components which can be further classified into following main groups of chemicals: aliphatics including alkanes, alkenes and cycloalkanes, aromatics which include monoaromatic and polyaromatic hydrocarbons, asphaltenes, resins and traces of metals (Petrov, 1987; Mullins et al, 2007; Marshall & Rodgers, 2004) . Creosote, a by-product of tar distillation, is composed approximately of 85% polycyclic aromatic hydrocarbons (PAHs), 10% phenolic compounds; and 5% N-, S-, and O-heterocyclics (Mueller et al, 1989). Among these groups of chemicals, PAHs are of foremost environmental concern due to their recalcitrance (Bamforth & Singleton, 2005; Cerniglia, 1993) and toxicity (Douben, 2003; Bispo et al, 1999; Harvey, 1996) . Monoaromatics, such as benzene, toluene, ethylbenzene, and xylenes (BTEX), are also some of the most common pollutants which are frequently encountered on contaminated sites. Similarly, alkanes pose a significant problem due to their inertness and viscosity (Head et al, 2006). Chain length of alkanes is the major determinant of their bioavailability and toxicity. Short chain n-alkanes act

as solvents for cellular membranes and fatty compounds and are directly toxic whereas, long chain n-alkanes have deleterious effect on the macro- and microflora by forming oil films.

1.2. Hydrocarbons as a food source for microbes

About a century ago, the ability of bacterial isolates to use aliphatic and aromatic hydrocarbons as sole carbon and energy sources was reported by Söhngen (1913) . Since then, many phyla of bacteria such as Proteobacteria (α-, β- and γ-), Actinobacteria, Firmicutes, Deinococcus-Thermus and Bacteriodetes, filamentous fungi and yeasts have been found to be involved in alkane degradation. In addition to being an important component of crude oil (20-50%), alkanes are also produced by plants, algae and bacteria (Hornafius et al. 1999; Seewald 2003 Widdel and Rabus 2001); that explains the presence of small amounts of alkanes in the pristine ecosystems. It has been observed that pristine ecosystems- both terrestrial and aquatic, contain some amount of hydrocarbon-degrading microorganisms, and their relative abundance is considerably increased in oil-polluted sites. Alkane degraders usually have a versatile metabolism and they can utilize a vast array of carbon sources in addition to alkanes. Most alkane degraders are facultative and they prefer simpler carbon sources before turning to alkanes but some obligate alkane degraders, also known as hydrocarbonoclastic bacteria, have also been reported from polluted environments. *Alcanivorax*, *Thalassolituus*, *Oleivorans* (Yakimov et al. 1998, 2004) and *Oleispira* are some genera of hydrocabonoclastic bacteria found to be involved in biodegradation of oil spills in several environments (McKew et al., 2007; Coulon et al., 2007).

PAHs are introduced in the environment via pyrolysis of biomass. It is well known that living organisms produce certain benzenoid compounds (Moshier and Chapman, 1973). Moreover, plants produce a vast array of aromatic compounds such as flavonoids, chromenes, dibenzofurans, cresols, xanthones, and lignin. The potential to degrade aromatic compounds is widespread in bacteria, and in the natural environment these bacteria contribute to the breakdown of aromatic compounds and to the global carbon cycle. Due to the aromatic ring structure, degradation pathways for different PAHs are much more diverse and complex compared to alkanes. While some microbes can completely mineralize few PAH compounds, most individual species do not harbour all the enzymes required for the whole pathway of degradation. In most scenarios, many different microbial groups with diverse enzyme systems are involved in the degradation of petroleum hydrocarbons. Elaborate knowledge of microbial populations capable of thriving on specific components of petroleum under various environmental conditions would therefore be highly desirable.

1.3. Rhizoremediation of petroleum hydrocarbon polluted soils

Owing to the catabolic potential and versatility of microorganisms (mainly bacteria and fungi), bioremediation is an effective way of cleaning up petroleum hydrocarbon contaminated soil. Success of a bioremediation regime depends, however, on several factors which might govern the colonization, niche preferences, survival, growth and reproduction of suitable microbes in contaminated soils. These factors include nutrient availability, temperature, pH, oxygen, moisture, toxicity and bioavailability of contaminants to name a few (Boopathy, 2000; Vidali,

2001). By gaining an understanding and accordingly manipulating these factors, effectiveness of bioremediation can vastly be improved. Phytoremediation, *i.e.* use of plants for remediation has been proposed and applied as an alternative strategy for the removal of organic pollutants (Macek et al, 2000; Susarla et al, 2002; Schnoor et al, 1995; White & Newman, 2011). Depending on the mechanism by which remediation of organics can be achieved, common phytoremediation strategies include:

- Phytovolatilization: plants take-up the contaminant from soil and after transforming it into a volatile compound, release it into the atmosphere (Arnold et al, 2007; Doucette et al, 2003).
- Phytoaccumulation/phytoextraction: organics or metals are removed by pollution-accumulating plants as the contaminants are concentrated in the harvestable parts of plants (Boonsaner et al, 2011).
- Phytodegradation/phytotransformation: enzymes derived from plants and/or their associated microbes degrade organic contaminants (Dec & Bollag, 1994; Newman et al, 1997).
- **Rhizoremediation**: Enhanced degradation of organics by microbes living in the rhizosphere of plants (Kuiper et al, 2004; Gerhardt et al, 2009; Leigh et al, 2007; Mackova et al, 2009).

Some of these mechanisms may also function simultaneously during the remediation of contaminants by plants.

Rhizoremediation has been shown to be an effective strategy for the clean-up of petroleum hydrocarbon contaminated soils (Ramos et al, 2010; Mackova et al, 2009; Mikkonen et al, 2011; Hong et al, 2011; Yateem et al, 2007). The importance of association of plant roots with soil microorganisms has been recognized since the early 20th century when Hiltner (1904) defined "rhizosphere" as the zone of soil where microbes are influenced by the root system. Apart from the physical influence of roots (enhanced aeration in soil, microbial attachment site, nutrient transport etc.), enhanced degradation of petroleum hydrocarbons in the rhizosphere can mostly be attributed to the root exudates. Root exudates can be generally categorized as organic polymers (from root debris and sloughing cells), low molecular weight carbohydrates, amino acids, organic acid anions and an array of secondary metabolites (Walker et al, 2003; Bais et al, 2006). Fig. 1 summarizes some of the important mechanisms by which root exudates can enhance the degradation of petroleum hydrocarbons in rhizosphere. These mechanisms are briefly described as follows (Martin et al, 2014):

A) Direct degradation: plant roots may secrete enzymes such as laccases and peroxidases which may catalyse the oxidation of some PAHs common in fuel mixtures (Gerhardt et al, 2009).

B) Co-metabolism, enzyme induction/selective enrichment: Many secondary metabolites in root exudates structurally resemble PAHs common in petroleum hydrocarbons. Hence, co-metabolism of PAHs along with the plant secondary metabolites is a major route of degradation of recalcitrant hydrocarbons (Cunningham et al, 1995; Fletcher & Hegde, 1995).

C) Increased bioavailability: It is well known that microorganisms can secrete biosurfactants which increase the solubility of certain organics and also favour the attachment

of microbes to oil droplets. Similarly, root exudates may also contain lipophilic compounds which promote the bioavailability of petroleum hydrocarbon compounds (Read et al, 2003).

D) Degradation via mycorrhizal symbionts: Extracellular enzymes secreted by mycorrhizal symbionts of plants can degrade some aliphatic and aromatic components of PHCs (Harms et al, 2011).

E) Increased biomass and/or activity- This is perhaps the most commonly mentioned role of root exudates, also referred to as the rhizosphere effect. Increased supply of nutrients and energy to microbes via root exudates leads to a higher abundance and activity of microbes in the rhizosphere.

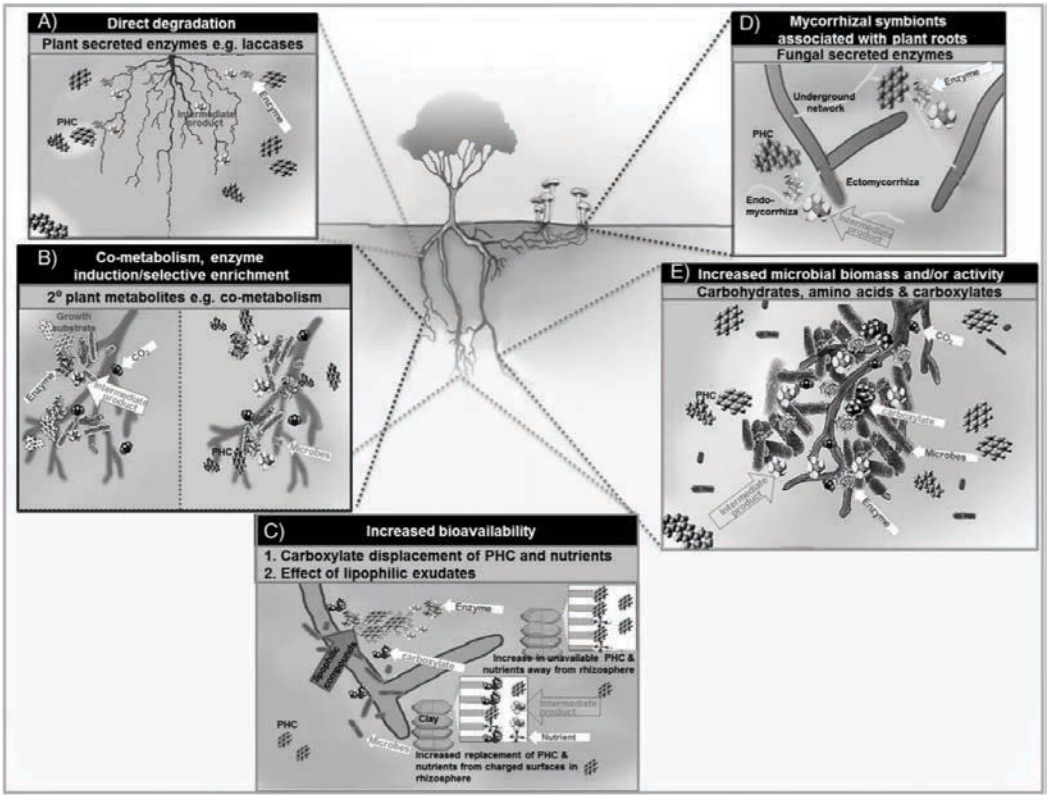


Fig. 1: Mechanisms by which root exudates can enhance biodegradation of petroleum hydrocarbons. Source: Martin et al., 2014

1.4. Poplars (aspen) for remediation of organics

Trees and perennial grasses are preferentially employed in phytoremediation as unlike annual plants, they do not need yearly replantation. Easy propagation, fast growth, phreatophytic root systems, high water uptake rate, high absorption surface areas and tolerance to contaminants are some of the important properties that make some tree species a practical choice for phytoremediation (Cook & Hesterberg, 2013). Among the most frequently studied tree species in context of phytoremediation of organic contaminants are poplars (Populus spp.), willows

(Salix spp.), red mulberry (Morus rubra) and birches (Betula spp.)(Vervaeke et al, 2003; de Carcer et al, 2007; Yrjälä et al, 2010; Sipilä et al, 2008; Tervahauta et al, 2009).

Populus is a member of the Salicaceae family, which also includes willows. In addition to around 30 species occurring in the northern hemisphere, *Populus spp*. also have the ability to cross within the genus, both in the wild and through controlled breeding, giving rise to a number of potential clones (Dickmann and Stuart 1983). Due to their ability to readily form hybrids, poplars have been crossed by foresters for years in order to maximize growth rates and yield (Klopfenstein et al, 1997). *Populus* has been considered a model woody plant due to its relatively modest genome size, rapid early growth, ease of clonal propagation, and routine transformation protocols (Bradshaw et al, 2000; Taylor, 2002; Wullschleger et al, 2002). Sequencing of the whole genome of *Populus trichocarpa* and availability of an extensive expressed sequence tag (EST) database has further added to the long list of importance research attributes of *Populus* species (Sterky et al, 2004; Tuskan et al, 2006). Owing to these advantages, hybrid poplars were originally bred and used as cash crops for pulp and energy. In past two decades, hybrid and transgenic poplars have extensively been utilized in the remediation of a broad range of organic contaminants (Table 1). It can be comprehended that poplars can remediate organic contaminants via various mechanisms including phytovolatilization, phytodegradation as well as rhizoremediation. This literature survey clearly indicates a knowledge gap regarding the structural and catabolic bacterial communities in polluted- Populus rhizosphere. Dynamics of rhizosphere associated bacterial communities in hybrid aspen (*Populus tremula* × *Populus tremuloides*) growing in hydrocarbon contaminated soils have been explored in this thesis project (**I &II**).

Table 1: Summary of studies involving poplars for remediation of organic contaminants

***Populus* species**	**Organic contaminant**	**Theme of study**	**Reference**
Populus deltoids × *nigra* DN34	Atrazine	Vegetative uptake and degradation of atrazine in the rhizosphere.	Burken & Schnoor, 1996
Populus deltoides × *nigra* DN-34, Imperial Carolina	Atrazine & benzene, toluene, and xylenes (BTX)	First report of enumeration of specific microbial populations (total heterotrophs, denitrifiers, pseudomonads, BTX degraders, and atrazine degraders) in *Populus* rhizosphere.	Jordahl et al, 1997
Populus trichocarpa × *P. deltoids*	Trichloroethylene (TCE)	A comprehensive study demonstrating efficient metabolism of TCE by axenic cultures of Populus cells, poplar cuttings and in a field trial with poplar trees.	Gordon et al, 1998
Populus deltoids × *nigra* DN-34	Benzene	Laboratory experiments investigating the toxicity response of poplar cuttings to benzene exposure, contaminant distribution in plant tissues, contaminant degradation in the soil profile, and contaminant volatilization from the soil and plant tissues.	Burken et al, 2001
Populus deltoides × *nigra* DN-34, Imperial Carolina	Methyl tert-butyl ether (MTBE)	A 3 phase study involving- A) a laboratory bioreactor study that examined the fate and transport of ^{14}C-radiolabeled MTBE in hybrid poplars B) mathematical modelling investigating the influence of deep-rooted trees on unsaturated and saturated groundwater flow & C) a field study at a site with MTBE-contaminated groundwater where hybrid poplar trees were planted.	Hong et al, 2001
Populus nigra	Diesel	A report on effect of diesel fuel on plant growth and on the rhizosphere microflora. T-RFLP fingerprinting of rhizosphere bacterial communities & isolation and identification of hydrocarbon-degrading strains was performed and these strains were checked for the presence of *alkB* gene.	Tesar et al, 2002
Populus deltoides × *Populus trichocarpa*	Volatile organic compounds (VOCs)	Phytoremediation field experiment demonstrating that hybrid poplar trees mitigate the migration of a groundwater plume of volatile organic compounds.	Hirsh et al, 2003
Populus deltoides × *nigra* DN-34	RDX	Poplar tissue cultures and leaf crude extracts were shown to mineralize RDX upon exposure to light.	Van Aken et al, 2004
P. trichocarpa × *deltoides* cv. Hoogvorst	Toluene	First report of *in planta* horizontal gene transfer among plant-associated endophytic bacteria. Inoculation of *Burkholderia cepacia* VM1468 containing pTOM-Bu61 plasmid (carrying genes for toluene degradation) had a positive effect on plant growth in the presence of toluene.	Taghavi et al, 2005

Populus trichocarpa × *deltoids* cv. "Hazendans" and "Hoogvorst"	BTEX	Endophytic bacteria were isolated from the root, stem and leaf of two cultivars of poplar tree growing on a site contaminated with BTEX compounds. They were further characterised genotypically (16S rRNA sequencing) and phenotypically by their tolerance to target pollutants. The endophytic bacteria exhibited marked spatial compartmentalisation within the plant.	Moore et al, 2006
Populus deltoides × *nigra* DN-34	RDX	First report showing that the exposure of poplar plants to RDX results in the induction of several genes that are involved in explosive detoxification such as glutathione S-transferases (GSTs), cytochrome P-450s (CYPs), NADPH-dependent reductases, and peroxidases.	Tanaka et al, 2007
Populus deltoides × nigra, DN34	Polychlorinate-d biphenyls (PCB)	Report on the transport of PCBs through whole plants designed for use in treatment at disposal facilities	Liu & Schnoor, 2008
Transgenic *Populus tremula* × *tremuloides* var. Etropole	2,4,6-trinitrotoluene (TNT)	Transgenic hybrid poplars expressing bacterial nitroreductase gene, *pnrA*, were shown to tolerate and take-up greater amounts of TNT	Van Dillewijn et al, 2008
Populus deltoids × (*Populus trichocarpa* × *Populus deltoides*) cv.Grimminge	TCE	First in situ inoculation of poplar trees, growing on a TCE-contaminated site, with the TCE-degrading strain Pseudomonas putida W619-TCE. P. putida W619-TCE was extablished and enriched *in planta* as a poplar root endophyte and by further horizontal gene transfer of TCE metabolic activity to members of the poplar's endogenous endophytic population.	Weyens et al, 2009
Populus tremula × *P. alba* and *P. trichocarpa*	Chlorpyrifos	Hydroponic study demonstrating the efficiency of poplars in uptake and translocation of Chloropyrifos.	Lee et al, 2012
Populus nigra (var. italica)	Petroleum hydrocarbons, PCBs and metals	Phytoremediation field study at a site historically contaminated with petroleum hydrocarbons, PCBs and metals; poplars together with horse manure treatment were shown to be effective in remediation. Biogeochemical parameters were monitored by SDS PAGE.	Doni et al, 2012
Populus deltoides × *Populus nigra*	Poly aromatic hydrocarbons (PAHs)	Efficacy of *Burkholderia fungorum* DBT1 (a strain isolated from oil refinery discharge and capable of degrading dibenzothiophene, phenanthrene, naphthalene, and fluorene) was demonstated as an endophyte in poplars during 18 weeks greenhouse experiment	Andreolli et al, 2013

1.5. Monitoring microbial populations in contaminated ecosystems using culture-independent tools

The "great plate count anomaly" implies that the readily cultured microbes are like the 'weeds' of microbial world and they represent less than 1% of the total microbial diversity (Staley & Konopka, 1985). This figure (1%) was however estimated by comparing plate counts with direct microscopic counts of microorganisms in environmental samples and has been criticized (Donachie et al, 2007). The pure culture based studies still retain their significance for the development and interpretation of molecular analyses in microbial ecology. Nevertheless, the tremendous amount of information generated by the application of molecular techniques has transformed our view of the microbial world (Hugenholtz, 2002). The key idea behind this approach was to explore the uncultured microbial diversity by retrieving and sequencing the macromolecules directly from environmental samples (Pace, 1997). While, the initial progress with this approach was mostly focused on the phylogenetic marker - 16S rRNA gene, techniques targeting specific functional marker genes and metagenomes have been developed and used over the recent decade (Handelsman, 2004; Albertsen et al, 2013; Van Elsas & Boersma, 2011; Hirsch et al, 2010). The inherent complexity of microbial communities has become quite evident. Culture-independent techniques are based on a holistic approach i.e. studying a natural habitat directly as opposed to a reductionist approach, i.e. studying each organism in isolation.

1.5.1. DNA based microbial community profiling

DNA based community profiling generally begins with the direct isolation of DNA from environmental samples followed by polymerase chain reaction (PCR) amplification of marker genes using universal primers capable of amplifying the target genes from a broad range of taxa (Nocker et al, 2007). The marker gene used could either be a phylogenetic marker like 16S rRNA gene or a functional gene encoding for a specific protein. The PCR products (amplicons) are further subjected to a suitable profiling method (Torsvik & Ovreas, 2002; Muyzer & Smalla, 1998; Liu et al, 1997; Suzuki et al, 1998; Fisher & Triplett, 1999; Dohrmann et al, 2004; Schwieger & Tebbe, 1998). Most profiling methods achieve the differentiation in the PCR products based on:

A) Location of restriction endonuclease sites: Terminal restriction fragment length polymorphism (T-RFLP), restriction fragment length polymorphism (RFLP), amplified ribosomal DNA restriction analysis (ARDRA, specific for 16S rRNA genes)

B) Melting behaviour of double stranded PCR products related to the sequence composition and primary structure of target gene fragments: Denaturing gradient gel electrophoresis (DGGE) and thermal gradient gel electrophoresis (TGGE)

C) Electrophoretic mobility of single stranded DNA fragments in non-denaturing gels: Single-strand conformational polymorphism (SSCP)

D) Variation in length of entire gene fragments: Length heterogeneity PCR (LH-PCR), automated ribosomal intergenic spacer analysis (ARISA).

The resolution of these profiling methods varies considerably and selection of a suitable method is important for the accurate interpretation of microbial community structure and dynamics.

An alternative approach is amplicon sequencing of target genes and comparing the sequence composition *in silico* for assessing the differences in community structure. Clone library sequencing has been the traditional method of choice for a long time and has helped us to describe a considerable fraction of structural and functional microbial diversity. Nonetheless, even the most extensive clone library based sequencings have underestimated microbial diversity due to under-sampling (Curtis & Sloan, 2005; Quince et al, 2008). In the last few years, next generation sequencing (NGS) platforms such as 454 pyrosequencing, Illumina, SOLiD, Ion Torrent, Ion proton, PacBio and Oxford Nanopore technologies have revolutionized the field of sequencing by the massive amount of data produced at reduced price and time. Due to the possibility of 'barcode –tagging' which enables massive screening of multiple samples, some of these next generation sequencing methods are becoming a potential way of microbial community profiling.

In this thesis, **T-RFLP (I) and 454 amplicon sequencing (II &III)** have been employed for the profiling of structural and functional microbial communities. Some technical aspects related to these two methods are discussed in the following paragraphs.

Terminal restriction fragment length polymorphism

T-RFLP is one of the most widely used fingerprinting methods in microbial ecology owing to its relative simplicity, great reproducibility and provision for robust *in silico* data analysis (Liu et al, 1997; Osborn et al, 2000; Schütte et al, 2008; Abdo et al, 2006; Dollhopf et al, 2001). It has frequently been used not only for analysing 16S rRNA genes (Sipilä et al, 2008; Mummey & Stahl, 2003; Yrjälä et al, 2010) but also functional genes; e.g. genes involved in hydrocarbon degradation such as alkane hydroxylases (Giebler et al, 2013; Schulz et al, 2012; Paisse et al, 2011), ring hydroxylating dioxygenases (Vitte et al, 2013) genes involved in nitrogen metabolism (Bannert et al, 2011; Wessén et al, 2011) and methane production (Juottonen et al, 2012; Yrjälä et al, 2011).

Briefly, the gene of interest is amplified using fluorescently labelled primers (either one or both the primers can be labelled); the resulting labelled amplicons are then digested with one or more restriction enzymes and the final step involves the determination of size and relative abundances of the fluorescently labelled terminal restriction fragment (T-RF) using an automated DNA sequencer. The differences in the length (size in base pairs) of T-RFs depict the difference in sequence composition of the target gene and thus ideally represent distinct microbial groups. It is also possible to identify specific T-RFs by cloning and sequencing the target gene from the same environmental sample. With a suitable data-analysis approach, T-RFLP can be applied to assess treatment- specific, temporal or spatial changes in microbial communities. One should bear in mind while designing or using T-RFLP protocols that this method shares the technical limitations associated with PCR; for instance, selection of appropriate primer pairs can be tricky as they should be specific but still have a broad coverage of multiple taxa. Additionally, partial digestion of amplicons can lead to a faulty representation of diversity.

454 pyrosequencing

The 454 pyrosequencing platform is a pioneering landmark in the field of next-generation sequencing because it advanced the field in unprecedented ways beyond three main bottlenecks i.e. library preparation, template preparation and sequencing (Margulies et al, 2005; Ronaghi et al, 1998; Ronaghi et al, 1996). The impact of 454 can be gauged on the basis of the fact that, the solutions it provided have since been and still continue to be emulated by successive next-generation sequencing platforms (Shendure & Ji, 2008; Rothberg & Leamon, 2008; Metzker, 2009). Since its arrival, 454 pyrosequencing has become a popular choice for studying the microbial diversity in diverse environments through 16S rRNA amplicon sequencing (Sogin et al, 2006; Roesch et al, 2007; Will et al, 2010; Leininger et al, 2006; Lauber et al, 2009), for metagenomic (Dinsdale et al, 2008; Edwards et al, 2006; Mackelprang et al, 2011) and metatranscriptomic (Poretsky et al, 2009; Gifford et al, 2010; Lehembre et al, 2013) studies as well as microbial genome sequencing (Paliwal et al, 2014; Goodison et al, 2013). Application of 454 sequencing has also added to our understanding of microbial community structure and dynamics in polluted ecosystems (Uhlik et al, 2012; Kotik et al, 2013; Dos Santos et al, 2011; Singleton et al, 2011).

An overview of 454 pyrosequencing workflow and summary of steps involved are presented in Fig. 2. The raw data generated from 454 pyrosequencing is a series of images which are normalized and converted into flowgrams (SFF files). SFF files are the starting point of sequence data analysis. One can either begin the pre-processing of 454 sequences with the flowgrams or use the Fasta files and proceed with the standard bioinformatics tools. Flow values in the SFF files, however, can provide additional information about sequence quality that is not available in the pure nucleotide sequences. Next step after pre-processing and denoising is usually alignment of sequences (unsupervised or alignment to reference databases); alignment is one of the most important steps in the process of quality screening of sequences. Next, removal of chimeric sequences is performed. Good quality sequences are then clustered into operational taxonomic units (OTUs) and microbial diversity and community structure are further analysed via downstream statistical analysis.

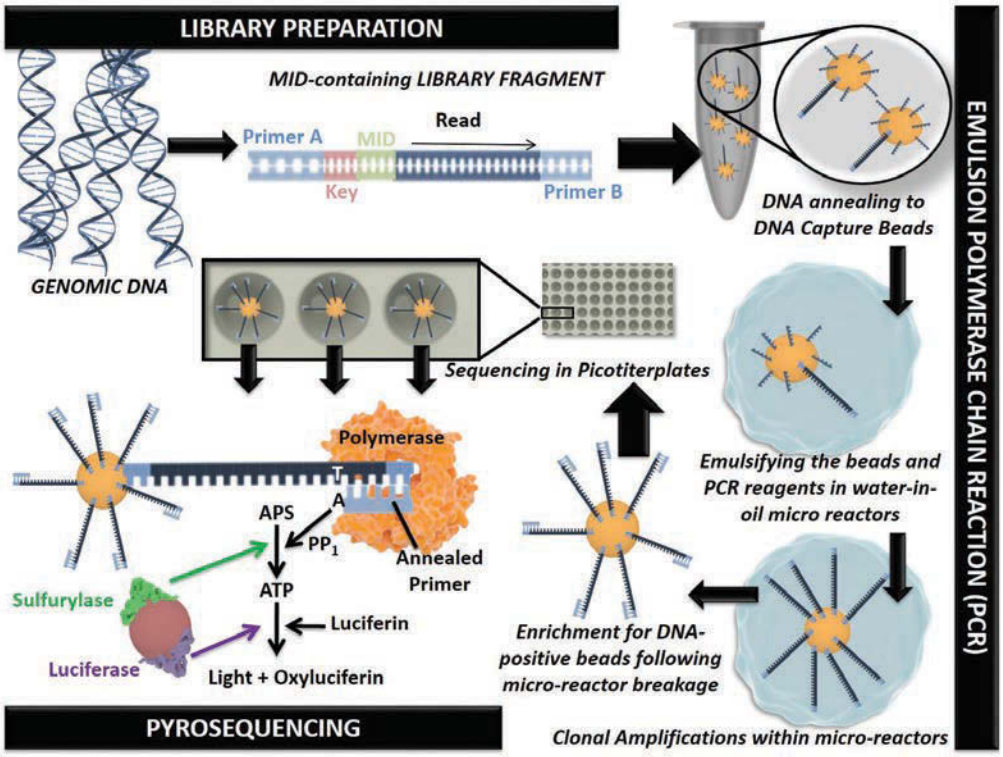


Fig. 2: A schematic illustration of the different steps in 454-sequencing. A) a library of amplicons/fragmented DNA or cDNA ligated to specific 5′- and 3′-end specific adapters is prepared B) followed by emulsion PCR wherein, each emulsion droplet behaves as a microreactor for PCR amplification, C) Beads with amplified fragments bound to them are released from emulsion and are loaded onto a picotiter plate with the capacity of one bead per well. Pyrosequencing takes place in the picotiter plates by a sequential flow of sequencing reagents across the plate, when a complementary nucleotide is added to a particular template in an extension reaction and a light signal is generated giving rise to flowgrams.(modified from Rothberg et.al. 2008)

1.5.2. Phylogenetic and functional/catabolic marker genes

Use of gene sequences as molecular clocks to decode phylogenetic relationships was first proposed by Pauling and Zuckerkandl (1965). The landmark study of Carl Woese and colleagues on the small subunit ribosomal RNA (16S rRNA and 18S rRNA) provided a framework for determining phylogenetic relationships among organisms (Woese, 1987; Woese et al, 1975). This pioneering work altered our view of evolution from a five kingdom to a three domain (Eukaryota, Bacteria and Archaea) paradigm (Woese & Fox, 1977; Woese, 1987; Woese et al, 1990). 16S rRNA gene has been established as the "gold standard" in microbial ecology studies and is by far the most commonly used phylogenetic marker due to its properties such as ubiquity, extreme sequence conservation, a domain structure with variable evolutionary rates and highly developed databases (Tringe & Hugenholtz, 2008). Nonetheless, existence of

multiple heterogeneous copies of the 16S rRNA gene within a single genome can hamper robust species differentiation (Crosby & Criddle, 2003). Alternative phylogenetic markers such as *rpoB*, *rpoD*, *recA* and *gyrB* have also been used (Case et al, 2007; Wang et al, 2007; Thompson et al, 2004). While 16S rRNA gene based community profiling gives a view of microbial community composition, it hardly can provide information on the community functions. This is due to the fact that phylogenetically unrelated groups of organisms can often carry out similar ecological processes and structural shifts in community does not always relate to corresponding functional differences. For this reason, genes coding for enzymes, which carry out specific metabolic processes, can give a better description of functional aspects of communities. In contaminated ecosystems, genes for catabolic enzymes involved in biodegradation pathways can be useful as functional markers for tracking the microbial groups capable of feeding on specific pollutants. Comparative and integrated analysis of phylogenetic and functional markers can possibly provide a better link between community structure and function.

1.5.3. Extradiol dioxygenase as a marker for aromatic degraders

As discussed in previous sections, PAHs are prominent environmental pollutants and therefore the enzymes involved in aromatic degradation pathways are of considerable interest. The aromatic ring system in these compounds makes them inert as substrates towards simple oxidation or reduction and thus requires elaborate degradation strategies. Given the vast variety in structure of aromatic compounds, it would take as many metabolic pathways as there are compounds for the microbes to make use of all of them if there was no coordination in their degradation. Most of the aromatic compounds are, however, channeled by **upper /peripheral pathways** into **lower/ central pathways** by their transformation to a few key central intermediates. Central/lower pathways then convert these intermediates into metabolites such as acetyl-CoA, succinyl-CoA and pyruvate (Fuchs et al, 2011). Aerobic and anaerobic environmental conditions further regulate the strategies of PAH degradation. The majority of reported bacterial aromatic degradation processes are aerobic but alternative strategies under low O_2 levels or even O_2-free conditions have also been identified (Gibson & S. Harwood, 2002; Heider & Fuchs, 1997; Meckenstock et al, 2004) .

Under aerobic conditions**, ring hydroxylating diooxygenases** oxidize aromatic rings by addition of both atoms of molecular oxygen to two adjacent carbon atoms to produce cis-dihydrodiols which are further acted upon by **cis-dihydrodiol dehydrogenases.** In the next step, **ring-cleavage dioxygenases** catalyze a reaction in which molecular oxygen is inserted into a C–C bond of dihydroxylated aromatic compounds, resulting in cleavage of the ring (Fig. 3). Two classes of **ring-cleavage dioxygenases** have been identified, on the basis of the mode of ring cleavage: **Intradiol dioxygenases** utilize non-haeme Fe III to cleave the aromatic nucleus *ortho* to (that is, between) the hydroxyl substituents, whereas **extradiol dioxygenases (EDO)** utilize non-haeme Fe II or other divalent metal ions to cleave the aromatic nucleus *meta* to (that is, next to one of) the hydroxyl substituents (Vaillancourt et al, 2006; Lipscomb, 2008). In catabolism of polyaromatic hydrocarbons (PAHs), the first aromatic ring structure is cleaved by the upper *meta*-pathway EDOs and the second ring by lower *meta*-pathway EDOs (Williams and Sayers 1994, van der Meer 1997, Lloyd-Jones et al. 1999).

In this thesis project, genes coding for upper *meta*-pathway **I.3.E group EDOs (**I.3.E designates type (I), family (3) and subfamily (E)**)** have been targeted as a marker to capture the functional diversity related to polyaromatic degradation in Bacteria. The target subfamily I.3.E contains enzymes for initial ring-cleavage of biphenyl, naphthalene and evidently also for polyaromates such as dibenzothiophene and phenanthrene, containing three aromatic rings (Denome et al., 1993; Pinyakong et al., 2003). The primer pair employed for amplification of extradiol dioxygenases in articles **I & II** was designed by Sipilä et al (2006). This primer pair theoretically amplifies a 469-bp fragment of the *nahC* gene of *Pseudomonas putida* Nah7 plasmid at position 131 to 600bp (Sipilä et al, 2006). The efficacy of these primers in successful amplification of extradiol dioxygenases from pristine and contaminated soils and PAH polluted birch rhizospheres has previously been demonstrated (Yrjälä et al, 2010b; Sipilä et al, 2006; Sipilä et al, 2008) . Broad specificity of these primers, *i.e.*, efficient amplification from Alpha-, Beta-, Gammaproteobacteria and gram positives such as *Rhodococcus* further makes them ideal for monitoring aromatic degraders in bioremediation and rhizoremediation studies.

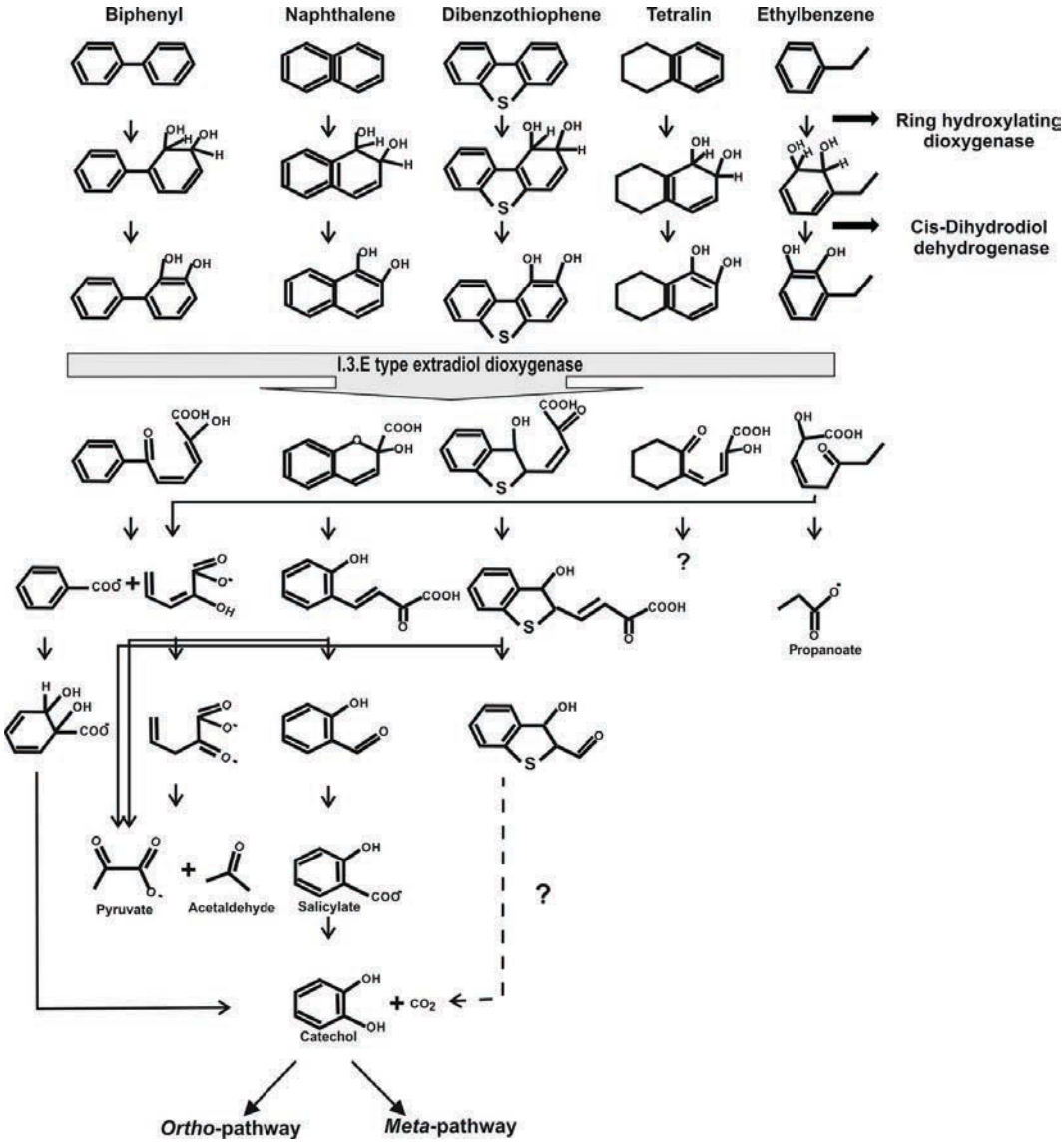


Fig. 3: Peripheral/ upper meta- pathway of degradation of aromatic compounds (biphenyl, naphthalene, dibenzothiophene, tetralin and ethylbenzene). 1.3 E type extradiol dioxygenase enzymes catalyze the third step in the degradation leading to cleavage of aromatic rings of these compounds. Adapted from Sipilä et al. 2009.

1.5.4. Alkane monooxygenase as a marker for aliphatic degraders

The ability to degrade alkanes under aerobic and anaerobic conditions is widespread among many phyla of Bacteria, some fungi and yeasts found in both pristine and polluted ecosystems (Rojo, 2009; Leahy & Colwell, 1990). Under aerobic conditions, different enzymes systems

initiate the biodegradation process by introducing oxygen in the alkanes based on their chain length (van Beilen & Funhoff, 2007). Different classes of these enzymes based on substrate range (alkane chain-length) have been listed in Table 2. It is noteworthy that the substrate ranges for these enzymes may overlap.

Among these classes of enzymes, the terminal oxidation pathway of alkanes involving **AlkB related alkane monooxygenases** has been most thoroughly studied (Rehm & Reiff, 1981; van Beilen et al, 2003; van Beilen & Funhoff, 2007). Fig. 4 represents the terminal pathway of alkane degradation and role of alkane monooxygenase as the key enzyme in the catabolism of medium chain length (C5- C16) alkanes. The functional Alk enzyme system comprises the transmembrane alkane monooxygenase AlkB (encoded by the *alkB* gene) and two co-factors named rubredoxin (AlkF) and rubredoxin reductase (AlkG). These co-factors are responsible for transferring the electrons involved in alkane hydroxylation by AlkB (van Beilen et al, 2001; van Beilen et al, 2003).The gene encoding for alkane monooxygenases (*alkB*) has been used as functional marker for tracking the abundance and diversity of alkane degrading bacteria in a variety of pristine (Schulz et al, 2012) and polluted soils (Bell et al, 2011; Pérez-de-Mora et al, 2011; Schloter, 2014; Whyte et al, 2002) and rhizosphere (Andria et al, 2009) . Primer pairs designed by Kloos et al (2006) were used for targeting the alkane degraders from oil polluted soil and rhizospheres in article **II** of this thesis. These primers were previously shown to successfully amplify *alkB* genes from Alpha-, Beta-, Gammaproteobacteria, Flavobacteria, Firmicutes and Bacilli from variety of soils (Jurelevicius et al, 2013).

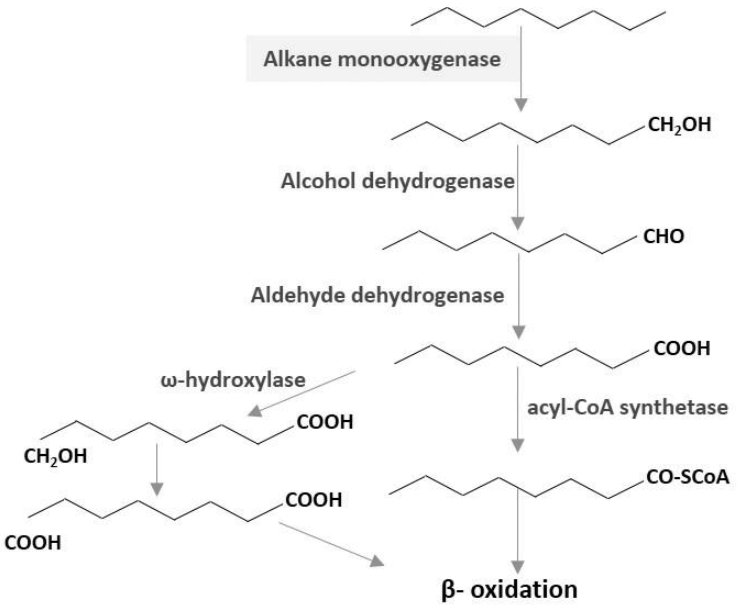


Fig. 4: Terminal oxidation pathway of alkane degradation. Alkane monooxygenase gene/enzyme studied in this thesis has been highlighted. Modified from Rojo, 2009

Table 2: Enzyme classes involved in aerobic degradation of alkanes

Substrate range	Enzyme class	Composition and cofactors	References
C1-C8 (halogenated)-alkanes, alkanes, alkenes, cycloalkanes	Soluble methane monooxygenase (sMMO)	α2β2γ2 hydroxylase; dinuclear iron reductase, [2Fe–2S], FAD, NADH regulatory subunit	Fox et al, 1990; Lipscomb, 1994; Oldenhuis et al, 1989; Vorobev et al, 2011
C1–C5 (halogenated)-alkanes, alkenes	Particulate methane monooxygenase (pMMO)	α3β3γ3 hydroxylase trimer composed of PmoA, PmoB, PmoC; mononuclear copper and dinuclear copper in PmoB	Rosenzweig, 2011; McDonald & Murrell, 1997; Semrau et al, 1995
C5–C16 alkanes, fatty acids, alkylbenzenes, cycloalkanes, etc	AlkB-related alkane monooxygenases	Membrane hydroxylase; dinuclear iron rubredoxin; mononuclear iron rubredoxin reductase, FAD, NADH	van Beilen & Funhoff, 2007; van Beilen et al, 2003
C5–C16 alkanes, (cyclo)-alkanes, alkylbenzenes, etc	Bacterial P450 oxygenase systems (CYP153, class I)	P450 oxygenase; P450 heme ferredoxin; iron–sulfur ferredoxin reductase, FAD, NADH	Funhoff et al, 2006; Bell et al, 2006; van Beilen et al, 2006; Bell et al, 2010
C10–C16 alkanes, fatty acids	Eukaryotic P450 (CYP52, Class II)	Microsomal oxygenase; P450 heme reductase; FAD, FMN, NADPH	Črešnar & Petrič, 2011; Yadav & Loper, 1999; Iida et al, 2000
C10 –C30 alkanes	Dioxygenase	Homodimer; copper, FAD	Maeng et al, 1996; Throne-Holst et al, 2007
C20 –C32 alkanes	AlmA type *n*-alkane monooxygenase	flavin-binding monooxygenase	Wang & Shao, 2012; Wentzel et al, 2007
C15 –C36	LadA type long-chain *n*-alkane monooxygenase	flavoprotein monooxygenase; NAD(P)H-dependent flavin reductase and monooxygenase	Li et al, 2008; Feng et al, 2007

1.6. Successional dynamics of microbes in polluted soils and rhizospheres

Comprehending the patterns of temporal changes in communities is one of the major pursuits in ecology. Ecological succession refers to a process of more or less deterministic developments in the composition or structure of an ecological community (Connell & Slatyer, 1977; Walker & Del Moral, 2003; Horn, 1974). In plant ecology, succession has typically been divided into two major categories: primary succession, which occurs on an un-colonized substrate (e.g. lava fields or glacial retreats) and secondary succession, which is triggered by a radical disturbance in a previously colonized environment (e.g. harvested agricultural fields or forest fire). Majority of our knowledge on the concepts and mechanisms of succession are derived from the field of plant ecology and microbial succession has received much less attention (Fierer et al, 2010). Recent advancements in culture-independent tools have helped microbial ecologists to overcome the methodological limitations in analysing the temporal variation of highly diverse and rapidly changing microbial communities in various ecosystems (Kowalchuk et al, 2000; Banning et al, 2011; Koenig et al, 2011; Wertz et al, 2007; Podell et al, 2014). Given the pivotal role of microbes in ecosystem functioning and the constantly increasing anthropogenic impact on ecosystem, predicting the responses and dynamics of microbial communities to different disturbance regimes has become crucial (Chapin III et al, 2000; Griffiths & Philippot, 2012). Studies on secondary succession can play an instrumental role in understanding the resistance and/or resilience of microbial communities in response to perturbations.

Petroleum hydrocarbon pollution is one of the major anthropogenically introduced disturbances in terrestrial and aquatic ecosystems. As discussed in the previous sections, some components of petroleum hydrocarbons are highly toxic to microbes and therefore will have a negative impact on microbial diversity. On the other hand, oil pollution also provides an increased nutrient input in form of hydrocarbons which can be degraded by various microbial groups. An event of oil pollution can thus trigger a series of successional changes in the structure and composition of microbial communities (Kaplan & Kitts, 2004; Dubinsky et al, 2013; Zhou et al, 2014; Roling et al, 2004). In rhizoremediation of petroleum hydrocarbons, the plant rhizosphere is an additional factor affecting the microbial community structure and can be expected to influence the successional trajectories of microbial communities. Furthermore, plant growth and even seasonal changes may have a considerable role in modifying the rhizosphere associated microbial communities (Smalla et al, 2001; Shakya et al, 2013; Dunfield & Germida, 2003). Taken together, these issues underpin the need for studies which can provide insight into microbial succession in polluted soils and rhizospheres. Secondary succession of bacterial communities in oil contaminated *Populus* rhizosphere was explored first in a greenhouse set-up (**I**) and then under field conditions (**II**) in this thesis.

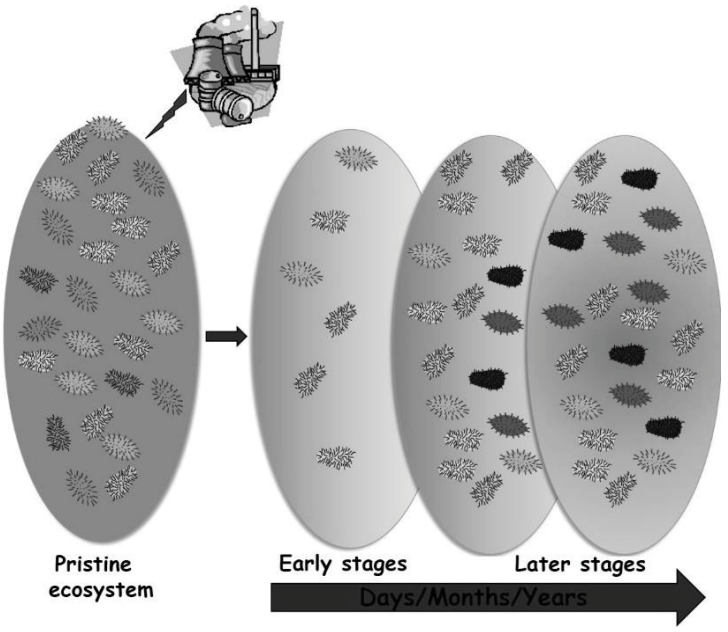


Fig. 5: A schematic representation of successional changes in microbial communities as an aftermath of pollution

1.7. Spatial heterogeneity in polluted sites and geostatistical approaches

Spatial heterogeneity of contaminants and other physicochemical and microbial properties in polluted sites are challenging for various steps of bio-/phytoremediation. Firstly, pre-evaluation or characterization of the extent of pollution and inherent catabolic potential of a site can be complicated as it requires an efficient sampling scheme to cover the spatial heterogeneity in polluted sites (Nedunuri et al, 2000). Secondly, for successful implementation of a suitable remediation regime at a field scale, *i.e.*, up-scaling of bio-/phytoremediation, it is important to consider the spatial patterns displayed by indigenous microbes which may co-vary with the spatial patterns of geo-chemical properties of a polluted site (Törneman et al, 2008; Bengtsson et al, 2010). Last but not least, evaluating the success of a remediation regime according to regulatory guidelines is often based on a point-by-point basis rather than an average of data points from across the site (Jennings & Petersen, 2006). All these issues highlight the necessity of integrating the spatial factor in study designs dealing with bioremediation.

Geostatistics is a powerful tool for gaining insights into such spatial heterogeneity. It includes a set of statistical methods for incorporating the spatial coordinates in data processing, modelling and description of spatial patterns, prediction at un-sampled locations and assessment of uncertainty related to these predictions (Goovaerts, 1998; Goovaerts, 1999). The techniques of geostatistics were initially developed and used in mining, petroleum industry and hydrogeology. Gradually, geostatistics paved its way from the spatial analysis of physicochemical soil properties to studies of spatial patterns of plants (Vieira, 1983; Sutherland

et al, 1991), soil microorganisms (Wollum & Cassel, 1984) and other soil surface organisms (Rossi et al, 1992). In recent years, geostatistics has aided the versatile toolbox of microbial ecologists in deciphering various ecological phenomena associated with spatial scales (Enwall et al, 2010; Wessén et al, 2011; Bru et al, 2010; Philippot et al, 2009; Krause et al, 2009). While geostatistical approaches are often used by environmental agencies and consultancies for the characterization of contaminants of sites, these methods are rarely employed for assessment of microbial community structure and activity on such sites. Geochemical properties that effect the distribution of subsurface microbes do not vary randomly but exhibit spatial continuity (Brockman & Murray, 1997). Some key concepts of geostatistics which are routinely used in description, modelling, prediction and mapping of spatial soil properties and have been applied in article **III** in this thesis are described briefly as follows (Source: (Ettema & Wardle, 2002; Goovaerts, 1998)).

Autocorrelation (spatial dependence): Autocorrelation is the statistical term for heterogeneity in spatial data. It quantifies the resemblance between neighbours as a function of spatial separation distance. When near neighbours are more similar than are far neighbours, data are autocorrelated, violating the assumption of data independence in standard parametric statistics.

Semivariogram and semivariogram modelling: Semivariograms are used in the first steps of spatial prediction. Semivariogram is a function that relates semivariance (or dissimilarity) of data points to the distance that separates them. Its graphic representation can be used to provide a picture of the spatial correlation of data points with their neighbours. Main components of semivariograms include range, sill and nugget (Fig. 6). The distance where the model first flattens out is known as the **range**. Range represents the extent or patch size of heterogeneity. Sample locations separated by distances closer than the range are spatially autocorrelated, whereas, locations farther apart than the range are not. The value that the semivariogram model attains at the range (the value on the y-axis) is called the **sill. Nugget** is indicated by the intercept closer to origin in y-axis; it represents the variance due to sampling error and/or spatial dependence at scales not explicitly sampled. Difference between sill and nugget i.e. **partial sill** indicates the spatially dependent predictability of the property being studied. Higher value of nugget relative to partial sill corresponds to higher "noise" or sampling errors.

Semivariogram modelling involves fitting a semivariogram curve to the empirical data collected. The goal with this step is to obtain the best fit and to incorporate the existing knowledge about the phenomenon being studied into the model; the selected model is further used in making predictions. A vast range of models are available, for example, Circular, Spherical, Tetraspherical, Pentaspherical, Exponential, Gaussian, Rational, Quadratic, Hole Effect, K-Bessel, J-Bessel, Stable etc.; each model is designed to fit different types of phenomena more accurately (Goovaerts, 1998).

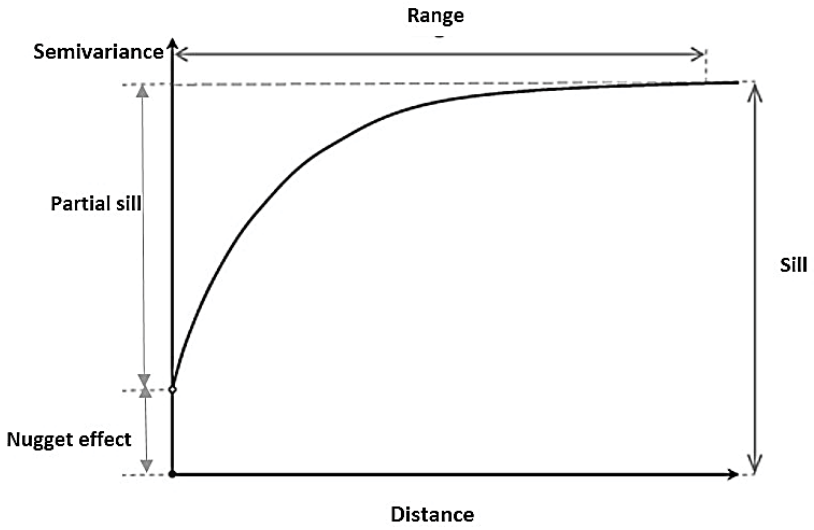


Fig. 6: A theoretical semivariogram

Spatial prediction, Kriging: Estimating soil properties at unsampled locations (spatial prediction) and mapping them are among the most important applications of geostatistics**.** Kriging is an interpolation technique in which the surrounding measured values are weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points. In this way, kriging serves two different purposes: quantifying the spatial structure of the data and producing a prediction. Kriging was named after the South-African mining engineer Danie G. Krige (Krige, 1976). There are many different kriging algorithms and selection of an applicable algorithm depends on trends in data under study. Examples of these algorithms include- ordinary, simple, universal, probability, indicator, factorial, disjunctive kriging, co-kriging and multivariate factorial kriging (Goovaerts, 1999; Goovaerts, 2001; Oliver & Webster, 1990; Stein, 1999). Selection of a suitable algorithm should be made based on the semivariogram properties and cross-validation before producing a prediction map by kriging. Fig. 7 illustrates some examples of semivariogram models and associated predictive maps.

Spatial scaling of biodiversity remains a central goal in ecology. For a long period, spatial variability in patterns exhibit by soil biota was considered as "noise" or complication in study design and interpretation. It is becoming evident, however, that spatial variability is a key rather than the obstacle in understanding the structure and function of soil biodiversity (Green et al, 2004). By combining molecular methods which provide a high resolution portrait of microbial populations and geostatistical approaches, it might be possible to unravel numerous obscurities regarding the generation and maintenance of soil microbial diversity.

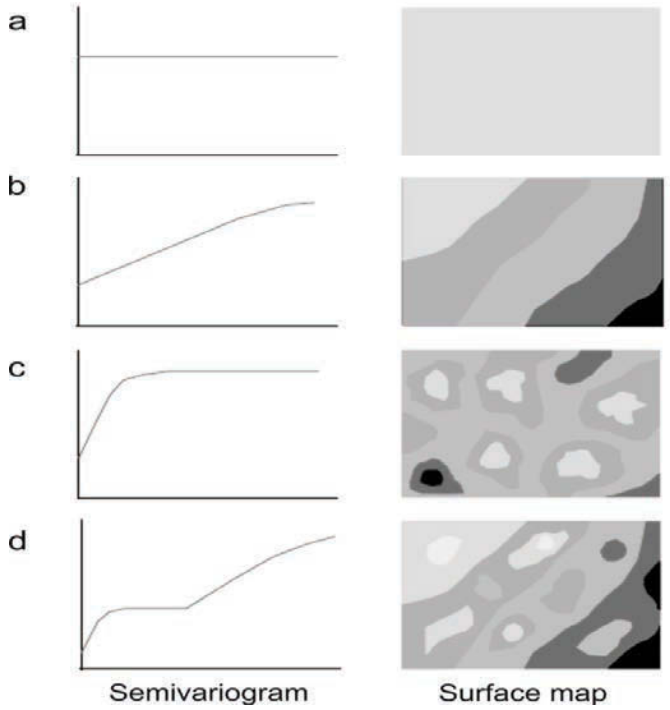


Fig. 7: Hypothetical semivariograms and associated prediction maps illustrating different patterns encountered during spatial analyses. A) Pure nugget effect, Variance is not spatially structured. B) Gradients and large scale heterogeneity. C) Small scale heterogeneity, patchy distribution of hotspots and cold spots. D) Nested heterogeneity of spatial patterns. Source- Ettema & Wardle, 2002

2. Outline and Aims

The main aim of this thesis project was to analyse and interpret temporal and spatial changes in microbial communities in hydrocarbon polluted soils and *Populus* rhizosphere in order to understand the ecological processes regulating these patterns and influencing rhizoremediation. This project was carried out on 3 different levels: a) a short term greenhouse study, b) a 2-year field study and c) a case study on an aged contaminated site (Fig. 8). The specific objectives were:

a) To determine the immediate and long term responses of soil and rhizosphere bacterial communities to hydrocarbon pollution **(I, II & III)**.
b) To explore the co-occurrence patterns of bacterial groups and to understand the possible ecological roles of co-occurring groups during the course of succession **(I &II).**
c) To access the catabolic diversity of bacteria in hydrocarbon contaminated ecosystem in order to understand the structure-function relationship of bacterial communities **(I &II).**
d) To evaluate the abiotic and biotic factors regulating the spatial distribution of microbial communities in an aged polluted site **(III).**

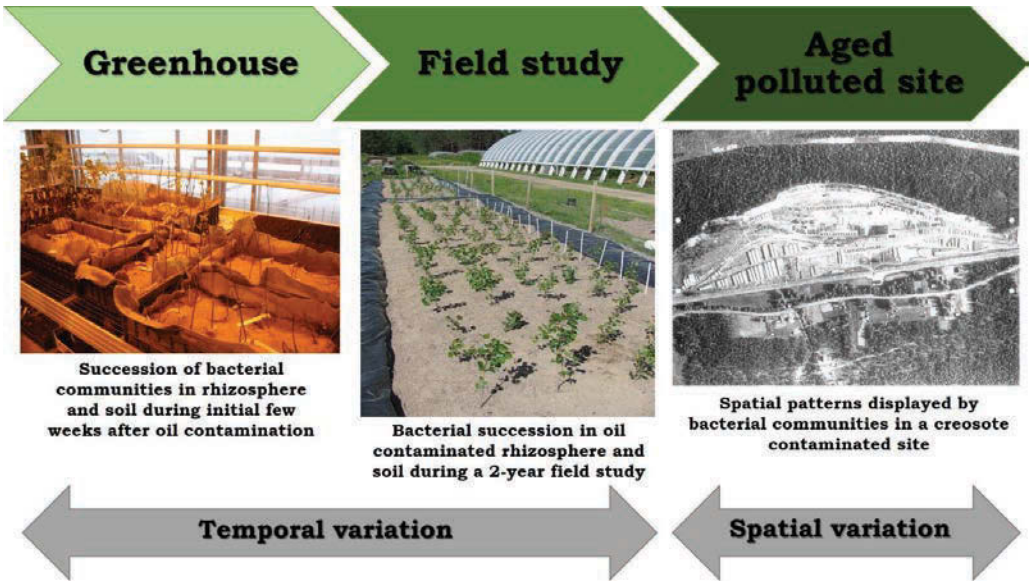


Fig. 8: Outline of the thesis project

3. Materials and methods

3.1.1. Study design and sample collection

A) Short term greenhouse experiment (I)

A 10-week greenhouse experiment was conducted in order to study the successional dynamics of bacterial communities in the *Populus* rhizosphere immediately after oil pollution. This study was conducted at the Finnish Forest Research Institute, Haapastensyrjä Tree Breeding Station in Southern Finland (60°37’24°25′, 1250 d.d.). Hybrid aspen (*Populus tremula* × *Populus tremuloides*) seedlings (20 cm minimum height) were grown in containers (40 × 60 × 12 cm, Takopot TA913, capacity 580 cm3, SCA Packaging, Finland) containing forest mineral soil. Top 25–30 cm of soil from which upper 3–5 cm was removed was used in this experiment. The soil was mixed with heavy fuel oil (Neste Oil ltd.) to an effective 1% concentration and incubated for a week at 18 °C before planting aspen. Microcosms were illuminated using high-pressure sodium lamps (400 W) and the incubation temperature in greenhouse was kept at 18 °C.

Composite soil samples (six sub samples), ~20 g, were taken from rhizosphere soil and bulk soil every week. The rhizosphere soil was obtained by lifting the seedling from the pot, shaking it gently and taking the sample from the remaining soil adhered to the roots. This experiment also included clean controls (same soil without oil addition). Altogether, this experiment included 6 different treatments: oil contaminated rhizosphere, oil contaminated bulk, and oil contaminated unplanted soil, clean rhizosphere, clean bulk and clean unplanted soil.

Fig. 9: Greenhouse study with hybrid aspen planted in oil contaminated (1%) soil.

B) Long term field study (II)

A field study was performed with hybrid aspen (*Populus tremula* × *Populus tremuloides*) planted in oil polluted soil; the sampling time points were planned to take into account the seasonal variation in the boreal climate zone (Fig 10). Soil used in this field study was excavated from an accidental oil spill site located by a motorway in Southern Finland. Following the excavation, the soil was transported to the Haapastensyrjä plant breeding station. After thorough mixing, the soil was planted with hybrid aspen seedlings (minimum 20 cm). In the same plot (3 m × 12 m), a small area (0.8 m× 2.5 m) was left unplanted and was partitioned from the rest of the plot to create an unplanted or bulk soil control. Samples from vegetated and vegetated treatments were taken at 7 time points during 2010-2012. Samples representing rhizosphere associated soil were collected by taking the 6 subsamples around one plant (10 cm below the ground and along 5-7 cm radius of the main root system); 2-3 replicate plants were sampled at each time point. The un-vegetated soil was sampled from the control box (10 cm below the ground and 2-3 replicates per time point).

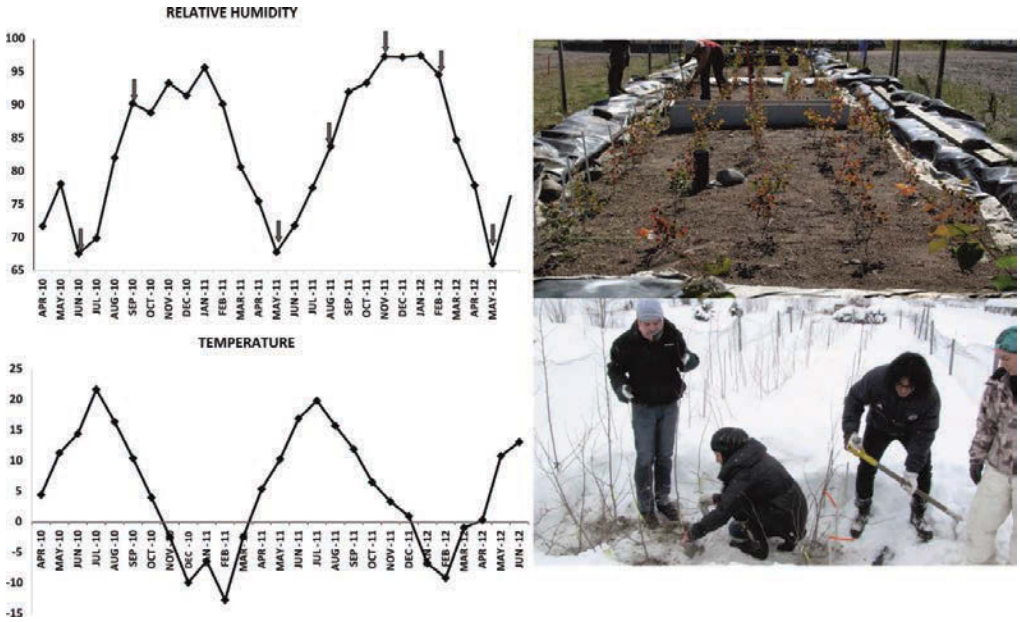


Fig. 10: 2-year field study with hybrid aspen planted in oil contaminated soil. Left panel represents the relative humidity and temperature variation during the study period and grey arrows in the relative humidity graph indicate the sampling time points.

C) Aged creosote-contaminated site (III)

An aged creosote-contaminated site located in South-eastern Finland (60° 55' 00" N and 27° 26' 00" E, Luumäki) was studied. This site was a wood preservation facility located in South-eastern Finland where creosote had been used to impregnate railroad sleepers between 1947 and 1958, resulting in high concentrations of creosote in the soil. As per the previous site investigations by Golder associates about ~ 1.3 hectares area in the central part, where the treatment facility was located, was characterized to be heavily contaminated. A sampling grid was created for the site (Fig. 11). The central area which was anticipated to be more contaminated was sampled using a 20 m x 20 m grid. Surrounding, less contaminated areas were investigated using a 40 m x 40 or 20 m x 80 m grid. In each plot, five 0.5 m deep sampling pits were excavated with a shovel. Top soil was collected from 0-0.1 m, and samples from five pits (four from the corners of the plot/square and one from centre) were combined to form composite samples representing each plot. Approximately 0.25 l of soil was collected from each pit and combined as 1 l composite sample in a gas tight plastic sampling bag. The composite sample was mixed well in the bag and then divided into portions for different analyses.

Fig. 11: Sampling design for studying the spatial patterns of microbial diversity and activity at Luumäki site. The black rectangular box in the middle represents the plots which were chosen for the microbial analysis. In addition, few plots outside the heavily contaminated area were also included in the study (marked by filled black squares). The hollow circles in each plot represent the sampling points.

3.2. Methods

The methods used in analysis of structural and functional microbial communities, microbial activity and soil chemical parameters have been described in the original articles (**I, II & III**) and are listed in the table below.

Method	Described and used in:
Molecular methods	
DNA extraction from rhizosphere and soil	**I, II & III**
PCR amplification of 16S rRNA, *edo* (I.3.E) and *alk*B genes	**I, II & III**
T-RFLP fingerprinting of 16S rRNA gene communities	**I**
T-RFLP fingerprinting of *edo* gene communities	**I**
Cloning	**I**
Sanger sequencing of 16S rRNA clones	**I**
454 pyrosequencing of 16S rRNA gene amplicons	**II & III**
454 pyrosequencing of *edo* gene amplicons	**II**
454 pyrosequencing of *alk*B gene amplicons	**II**
Microbial activity assays	
Most probable number (MPN) count for hydrocarbon degraders	**I**
Basal respiration assay	**III**
Fluorescein diacetate (FDA) hydrolysis assay	**III**
Soil chemical analyses	
Macro- and micro nutrients	**II & III**
Total petroleum hydrocarbons (TPH)	**II & III**
16 USEPA priority PAHs	**III**
Heavy metals	**II & III**
pH	**I, II & III**
Sequence analyses (methods and software)	
T-RFLP electropherograms - GeneScan (Applied Biosystems)	**I**
Quality check of clone library sequences – Genedoc	**I**
In silico restriction digestion – Sequence manipulation suit	**I**
Sequence alignment – Clustal W	**I, II**
Basic local alignment search tool (BLAST)	**I, II, III**
Phylogenetic analysis - MEGA	**I, II**
Phylogenetic trees - iTOL	**II**
16S rRNA pyrosequencing data analysis - MOTHUR	**II & III**
edo and *alkB* pyrosequencing data analysis- MOTHUR	**II**
Distance matrix of translated functional genes sequences -PHYLIP	**II**
Statistical analyses (methods and software)	
Multivariate analysis (PCA, RDA, NM-MDS, NPMANOVA, Biplot analysis) – PAST & R statistical package	**I, II & III**
Hierarchical clustering – Cluster 3 & Java tree view	**I & II**
Diversity indices - PAST	**I, II & III**
Correlation – Graphpad Prism	**III**
Geostatistical modelling - ArcGIS	**III**
Spatial prediction (Kriging) - ArcGIS	**III**

4. Results and Discussion

This chapter provides an overview and discussion on key findings of the thesis. The detailed description with figures and tables, referred to in the following text, is presented in the original articles (**I, II & III**) of this thesis.

4.1. Secondary succession of bacterial communities in oil pollution (I & II)

Oil pollution causes a disturbance in the chemical and physical soil environment thereby altering the ecological niches available to soil biota. Soil microbes have the capacity to utilize hydrocarbons as a substrate. Nevertheless, depending on the level of pollution (concentration), oil compounds are also toxic to microbes. Followed by oil pollution, a series of changes (secondary succession) can be triggered in the microbial community structure. We studied the successional trajectories of bacterial communities in oil polluted soil planted with aspen in two different studies: i) a short-term greenhouse study and ii) a long-term field study.

4.1.1. Temporal patterns of bacterial communities during a 10-week greenhouse experiment (I)

We performed 16S rRNA T-RFLP fingerprinting of bacterial communities from contaminated and uncontaminated rhizosphere, bulk and unplanted soil samples collected weekly. A sudden decrease in soil bacterial diversity (2.2 to 1.4; Shannon indices) was noticed one week after the oil amendment. In the following weeks, an increase in diversity was observed in the oil rhizosphere, bulk and oil control (unplanted) but oil rhizosphere maintained a higher diversity compared to the other treatments (supplementary table 1**, I**). Most probable number (MPN) assay revealed an increase in the abundance of culturable oil-degraders in the polluted rhizosphere during the first 2 weeks (Fig 6, **I**). Initial decrease in diversity reflected the immediate disturbance caused by oil contamination and its deleterious effect on bacterial communities, which is typical in secondary succession (Horn, 1974). In the later weeks, however, the bacterial communities started to adapt to the oil contamination as noticed from the increasing diversity. Higher diversity of bacterial communities along with the increased number of oil degraders in rhizosphere samples indicated a rhizosphere effect introduced by aspen plants in oil contaminated soil.

Multivariate analysis (PCA) of bacterial community structure (Fig. 12) further indicated a clear effect of oil pollution on the bacterial communities as samples from oil contaminated treatments were separated from uncontaminated ones along PC1 (45.6% variance). A non-random or directional pattern of temporal development was observed in oil-polluted rhizosphere, bulk and unplanted bacterial communities during the course of study. Also, the temporal patterns observed in the contaminated treatments were alike, with oil bulk and oil control showing similar community structure during the time course. In contrast, bacterial communities in uncontaminated treatments showed random temporal dynamics with much less variation with time compared to the contaminated treatments. These results clearly indicated the successional changes initiated by oil pollution. While slight differences between rhizosphere, bulk and unplanted controls could be detected among the contaminated treatments, effect of oil pollution on bacterial community structure was more pronounced than the

rhizosphere effect. A recent study also reported similar findings; they noticed a lower impact of willow rhizosphere compared to hydrocarbon contamination levels on microbial community structure (Yergeau et al, 2014).

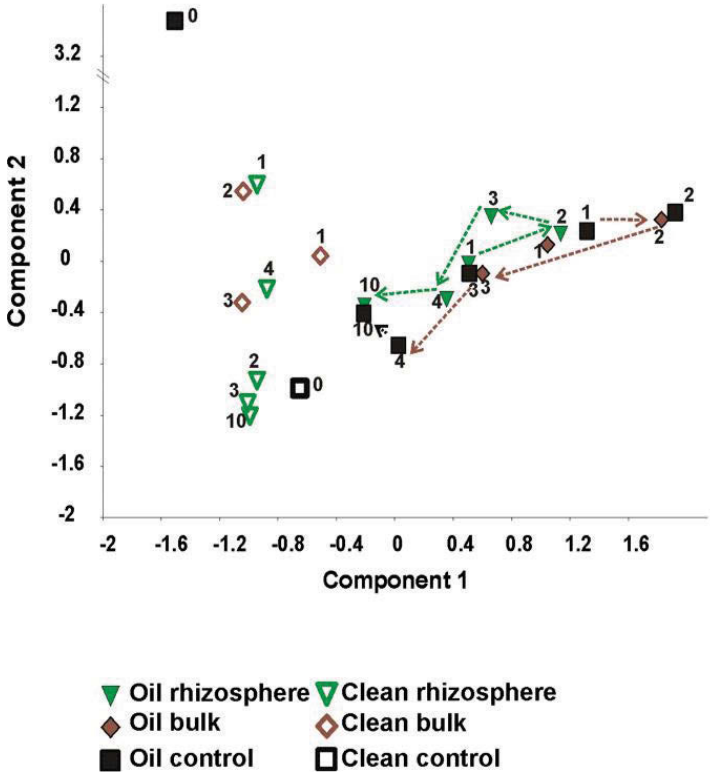


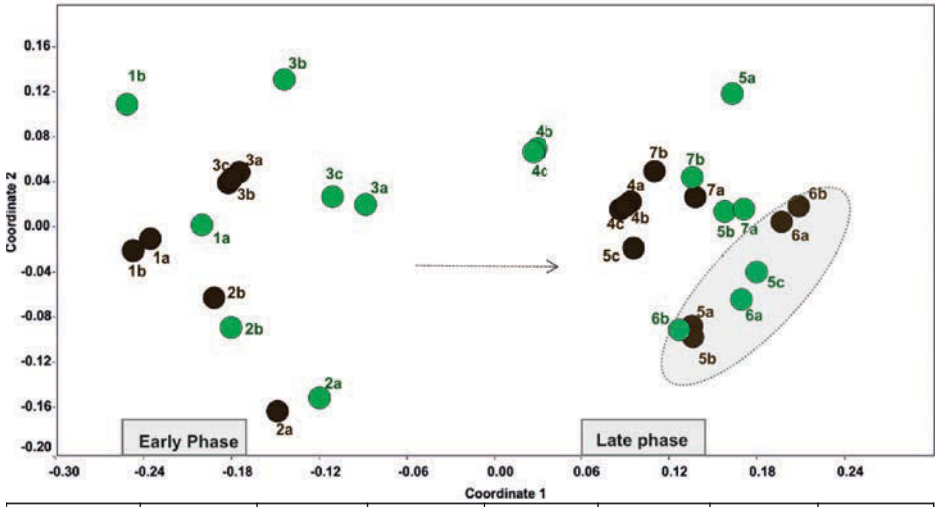
Fig. 12: Principal component analysis based on 16S rRNA gene T-RFLP fingerprinting of bacterial communities. Numbers in the sample code represent the week number when sampling was performed. Colour codes below the figure represent different treatments in this study.

4.1.2. Temporal patterns of bacterial communities during a 2-year field study (II)

We further aimed to analyse the long-term successional patterns of bacterial communities in oil-polluted aspen rhizosphere and soil in a field experiment. To this end, we sampled rhizosphere-associated and un-vegetated soil at 7 different time-points during 2 years; the sampling time-points were chosen to take into account the seasonal variations. 454 pyrosequencing of 16S rRNA gene amplicons was performed to assess the bacterial community dynamics. A total of 3215 OTUs (97% sequence similarity) were detected from 34 samples after removing the singletons. Two distinct clusters of bacterial communities (NPMANOVA; p=0.0001) were separated in multivariate analysis of 16S rRNA OTUs, referred to as early and late phase in the following text and figures (Fig. 13). Bacterial communities in contaminated rhizosphere and bulk (un-vegetated) soil from June 2010, September 2010 and May 2011

grouped in the early phase, whereas those from August 2011 to February 2012 grouped in the late phase cluster. While differences in rhizosphere and bulk communities from the same time points could be observed, sampling time points had a more prominent effect on bacterial community structure. Biological replicates from rhizosphere displayed a higher variation compared to biological replicates of bulk soil. During the late phase, bacterial communities in winter months (indicated as a grey ellipse in Fig. 13) showed marked variation (NPMANOVA p=0.0005) from spring and summer communities.

It has earlier been reported that during the course of oil degradation, due to differential use of hydrocarbon components, distinct successional changes occur in soil bacterial communities (Kaplan & Kitts, 2004; Röling et al, 2002; Yu et al, 2011; Roling et al, 2004) . We were able to notice clear successional changes in bacterial communities in oil polluted *Populus* rhizosphere and soil during the 2-year field study. We could not notice significant differences between rhizosphere and bulk diversity indices. A rise in diversity in both rhizosphere and bulk could be noticed at the second sampling time point. These results are similar to our short-term greenhouse study where the bacterial diversity increased noticeably after the initial drop indicating resilience of bacterial communities (**I**). Another similar observation in both these studies of secondary succession was a relatively higher effect of oil pollution compared to the rhizosphere effect on the bacterial community structure. Additionally, in this 2-year study, seasonal variation was shown to play a role in regulating the bacterial community structure during succession.



Sample code	1a, 1b	2a, 2b	3a, 3b, 3c	4a, 4b, 4c	5a, 5b, 5c	6a, 6b	7a, 7b
time-point	June 2010	Sep 2010	May 2011	Aug 2011	Nov 2011	Feb 2012	May 2012

Fig. 13: Non-metric multidimensional scaling (NM-MDS) based on Bray-Curtis distances of 16S rRNA OTUs (97% sequence similarity) from oil polluted rhizosphere and soil. Sample codes are mentioned below the NM-MDS plot. In the sample codes, a, b &c represent biological replicates. Rhizosphere samples are represented by green filled circles and un-vegetated soil samples are represented by brown filled circles.

4.1.3. Co-occurrence patterns of bacterial groups during succession (I&II)

We explored co-occurrence patterns of OTUs during the successional processes in both of our rhizoremediation studies (greenhouse and field experiment). Non-random co-occurrence patterns have been useful in unravelling the ecological processes governing community assembly rules in macro- and micro-ecology (Ulrich & J Gotelli, 2007; Gotelli & McCabe, 2002; Ruan et al, 2006; Barberán et al, 2012; Horner-Devine et al, 2007). It has also been suggested that analysing co-occurrence patterns varying with environmental parameters could give an insight into community structure-function relationships (Fuhrman, 2009). Monitoring the behaviour of co-occurring groups in relation to specific environmental factors such as pollution, influence of rhizosphere and time helped us to infer their possible ecological roles. Hierarchical clustering is a traditional approach in ecological studies to reveal relationships between communities sampled from different environments. We performed the hierarchical clustering of OTUs (instead of communities) based on their co-occurrence patterns across the whole time-series and in different treatments.

In our short term greenhouse study, we noticed two different linkage groups or clusters of 16S rRNA gene OTUs (T-RFs) (Fig. 4, **I**). Cluster A was composed of a group of bacteria present in all treatments: clean and oil-contaminated rhizosphere, bulk and unplanted soil, at all time points throughout the study but relatively more dominant in the clean treatments. Based on the *in silico* digestion of clone library sequences, some of the T-RFs were identified as Alphaproteobacteria (T-RF 65 and 152), Acidobacteria Gp I (T-RF 95 and 268) and Bacteroidetes (T-RF 92). These groups probably represented the generalists or k-strategists in oil pollution. While Alphaproteobacteria and Bacteroidetes contain some well-known degraders of oil components, Acidobacteria possibly represented the oil-tolerant groups in this study. Cluster B consisted of bacterial groups which were highly dominant in the oil-contaminated treatments and could hardly be observed in the clean treatments; most of the T-RFs belonging to this cluster were identified as Betaproteobacteria. This cluster represented specialists or oil degraders in oil-polluted soil and rhizosphere.

Next, we analysed co-occurrence patterns of bacterial groups in our long-term field study. 454 pyrosequencing of 16S rRNA gene amplicons used in this study provided a higher resolution and an improved identification of OTUs. Hierarchical clustering of 16S rRNA gene OTUs brought forward two major co-occurring groups: cluster A dominating in the early phase of succession and cluster B dominating in the late phase of succession (Fig 4, **II**); see section 4.1.2 for the clarification of early and late phases. The identified OTUs in early phase cluster belonged to Actinomycetales, Sphingomonadales, Rhizobiales, Caulobacterales, Burkholderiales and Xanthomonadales. The late phase communities were dominated by OTUs affiliated to unclassified Gammaproteobacteria, unclassified Betaproteobacteria, Rhizobiales, Acidobacteria GpI, Acidobacteria Gp4 and unclassified bacteria. Subcluster 3, which was a part of the late phase cluster represent a group of bacteria which were also observed in the early phase rhizosphere communities. Members of subcluster 3 belonged to Sphingomonadales (*Parvibaculum*), unclassified Beta and Gammaproteobacteria and Acidobacteria Gp4. Apart from these early and late phase members, two OTUs affiliated to the order Rhizobiales could be detected all along the study duration. This analysis enabled us to identify the bacterial groups dominating in the early and late phase of secondary succession. By comparing the co-occurrence patterns of 16S rRNA gene OTUs with those displayed by catabolic gene OTUs (alkane hydroxylases and extradiol dioxygenase), we were further able to infer the possible

catabolic functions of community members at different successional stages. Community dynamics and co-occurrence patterns of catabolic communities are discussed in the following sections.

4.1.4. Catabolic genes in oil pollution (I&II)

Two different marker genes were used for studying the bacterial catabolic diversity in oil-polluted rhizospheres and soils: extradiol dioxygenase as a marker for degraders of aromatics and alkane monooxygenase as a marker for degraders of aliphatics.

A. Extradiol dioxygenases: community dynamics and diversity (I&II)

Extradiol dioxygenases catalyse the *meta*-cleavage of aromatic ring as the third step in the peripheral degradation pathway of aromatic compounds. These enzymes have been characterized in Alpha-, Beta-, Gammaproteobacteria and Actinomycetales and have been shown to degrade a broad range of aromatics including biphenyls, naphthalene, phenanthrene, anthracene, fluoranthene, chrysene, naphthalenesulfonates, 6-dimethyldibenzothiophene tetralin, dibenzothiophene and ethylbenzene (Pinyakong et al, 2003; Andujar et al, 2000; Story et al, 2001; Lu et al, 2000; Keck et al, 2006; Jeon et al, 2006; McLeod et al, 2006).

T-RFLP fingerprinting of *edo* gene communities was performed to monitor the population of aromatic degraders in our greenhouse study (**I**). Marked difference in *edo* gene communities of contaminated and clean treatments was observed (Fig 1b, **I**). Distinct temporal variations were detected in the oil contaminated treatments and also in the clean rhizosphere. Temporal variations of extradiol dioxygenase communities in the clean rhizosphere could be due to the secretion of secondary metabolites which resemble some aromatics, thereby stimulating the aromatic degraders. Hierarchical clustering of T-RFs revealed 3 important co-occurring clusters. These T-RFs clustered together based on their similar temporal dynamics and occurrence patterns in different treatments (Fig. 5, **I**). Cluster X (T-RF 62, 64, 67, and 69 bp) represented the T-RFs dominant in the clean treatments and rarely present in the contaminated treatments. Cluster Y (56, 53, 77, 73, and 80 bp) represented T-RFs present in both the clean and contaminated treatments. Cluster Z (330, 319, 384, 465, 395, 461, and 463 bp) consisted of the T-RFs clearly dominant in the polluted treatments. Co-occurrence patterns of extradiol dioxygenase OTUs were comparable to the 16S rRNA genes in this short-term greenhouse study.

We performed 454 pyrosequencing of extradiol dioxygenase amplicons from 34 samples taken at 7 time points from rhizosphere and un-vegetated soil during the 2-year study period (**II**). A total of 22391 good quality sequences were clustered into 153 EDO OTUs (90% amino acid sequence similarity). Similar to the 16S rRNA communities, a clear succession was displayed by EDO communities (Fig. 3a, Article II). Differences in rhizosphere and bulk communities were more prominent in the late phase of succession. We also observed a steady rise in EDO diversity (Shannon indices, supplementary figure 3, **II**) during the course of succession.

Co-occurrence patterns of EDO OTUs were determined by hierarchical clustering to differentiate between the members of late phase and early phase communities (Fig. 5, **II**). In order to identify the putative hosts of the extradiol dioxygenase sequences detected in this study, we constructed a phylogenetic tree with the OTU representative sequences; closest hits

from NCBI database (-nr) and sequences from different clusters of the EDO tree from a study by Sipilä *et al.* 2008 were included as reference sequences. Members of early phase communities grouped in 2 different clusters of EDO phylogenetic tree (Fig 6, **II**). The cluster harbouring some of the most abundant OTUs of early phase had representative sequences from various genera of Sphingomonadales; these group of enzymes, characterized in *Sphingomonas* species, have been reported to target a diverse set of polyaromatic compounds (Jouanneau & Meyer, 2006). Other members of early phase community belonged to a phylogenetic cluster with no cultured representatives.

In the late phase of succession, community composition of EDOs was much more diverse and the sequence dissimilarity between the members was higher according to the EDO tree. OTUs dominating in late phase showed sequence similarity to *Sphingomonas, Sphingopyxis, Novosphingobium, Burkholderia, Cycloclasticus, Polaromnas, Ralstonia, Pseudomonas and Rhodococcus* type extradiol dioxygenases. In addition to these cultured representative from which extradiol dioxygenase enzymes have already been characterized, we also found high sequence similarity of OTUs in our study with sequences from Alphaproteobacterial species *Parvibaculum* and *Sandarakinorhabdus.* However, neither the EDO enzymes have been purified and characterized from these strains, nor their degradative capacities towards PAHs have been tested. *Parvibaculum lavamentivorans* has been shown to play a role in the degradation of surfactants such as alkylbenzenesulfonates (Schleheck et al, 2004; Dong et al, 2004; Schleheck et al, 2011). Sequence from *Sandarakinorhabdus sp.* Strain AAP62 was in the same phylogenetic cluster as *Parvibaculum* and the genome of this photoheterotrophic fresh water strain has recently been sequenced (Zeng et al, 2013). High sequence similarity with the extradiol dioxygenase sequences from our study indicates the presence of these genes in these two strains; characterization of degradative capacities of these strains and efforts for isolation of respective enzymes would therefore be useful. Many clusters with no representatives from cultured strains were observed in our EDO phylogenetic tree (Fig. 14); these clusters indicate the tremendous unexplored diversity of these enzymes in contaminated soil and rhizosphere.

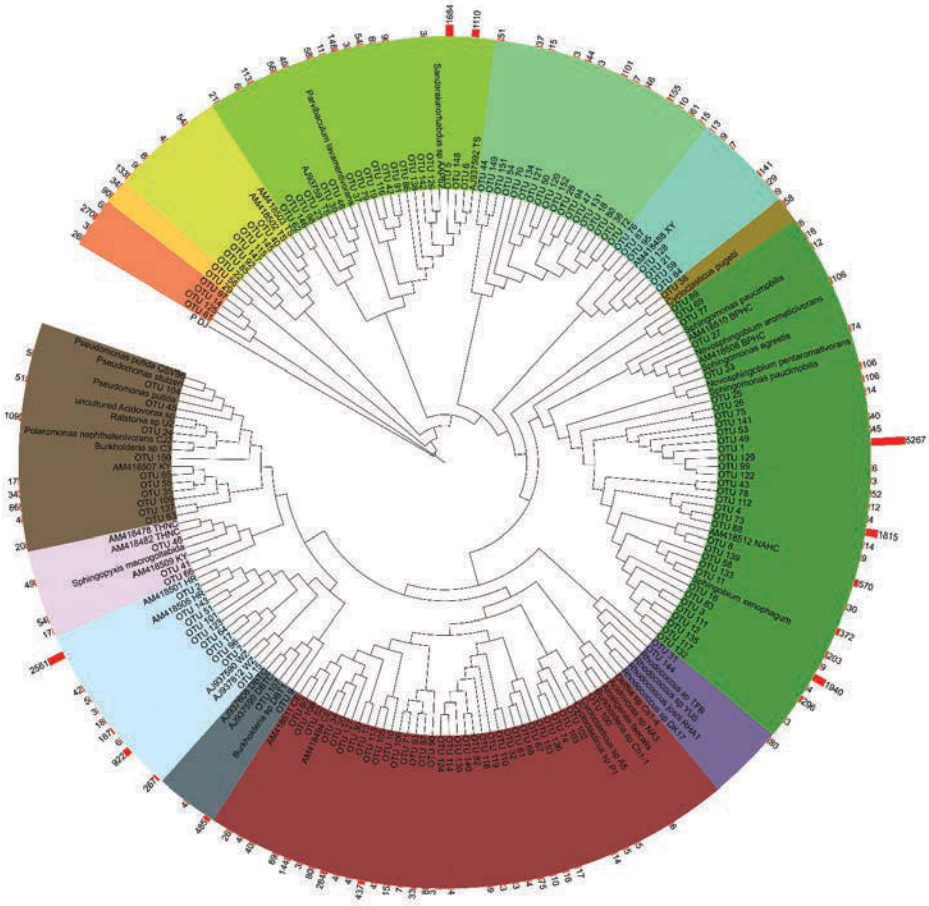


Fig. 14: Maximum likelihood tree of extradiol dioxygenase (EDO) amino-acid sequences based on the JTT model (Jones et al, 1992) . The bootstrap consensus tree was inferred from 500 replicates. One representative sequence from each EDO OTUs was included in the tree along with the closest hits from NCBI non-redundant protein database. OTUs with more than 2 sequences were chosen for better visualization. The red bars indicate the abundance of OTUs (number of sequences).

B. Alkane monooxygenases: community dynamics and diversity (II)

454 pyrosequencing of alkane monooxygenase amplicons was performed in order to reveal the community composition and dynamics of alkane degrading bacterial populations in the long-term field study. Primers used in this study are known to have a broad coverage of bacterial phyla making them suitable for community ecology studies (Jurelevicius et al, 2013; Kloos et al, 2006). A total of 53418 good quality sequences were obtained from 34 samples. After removing the singletons, 603 AlkB OTUs (95% amino-acid sequence similarity) were detected. Similar to the 16S rRNA communities, multivariate analysis (PCA) revealed the separation of alkane monooxygenase communities into early and late phases thereby showing a clear successional change. In the early phase, AlkB communities in rhizosphere showed marked differences from the bulk/unplanted communities (NPMANOVA p=0.0003). Hierarchical clustering of top 50 AlkB OTUs revealed three major co-occurrence patterns based on their

temporal behaviour (Fig 7, **III**). Cluster 1 was composed of OTUs which dominated in the early phase bulk and rhizosphere and thereafter underwent a more prominent decline in the late phase rhizosphere. Cluster 2 consisted of OTUs dominating in the late phase and Cluster 3 was composed of OTUs which could be observed throughout the study with some variation in relative abundances. These patterns were comparable to those displayed by the structural communities (16S rRNA). We noticed a considerable rise in diversity and decrease in dominance in AlkB communities during the course of succession (supplementary Fig 3, **II**).

AlkB communities in oil polluted *Populus* rhizosphere and soil were highly diverse and belonged to a broad range of taxonomic groups including Alpha-, Beta-, Gammaproteobacteria, Actinobacteria and Bacteroidetes (Fig. 15). The databases of alkane monooxygenases are far more extensive than extradiol dioxygenases thereby providing a better identification of hosts. Interestingly, the phylogeny of AlkB sequences did not correspond to the taxonomical closeness of the bacterial groups harbouring them. With an exception of few clusters in the AlkB tree, which were dominated by either Proteobacteria or Actinobacteria, sequences from different classes and phyla grouped together in the same clusters. These results were corroborated by a recent study by Giebler et al (2013b) ; they found that *alkB* and 16S rRNA gene derived phylogenies are not always congruent. This is possibly due the extensive horizontal gene transfer (HGT) of alkane monooxygenases among taxonomically distant groups which has been pointed out earlier (van Beilen et al, 2001; Heiss-Blanquet et al, 2005; Van Beilen et al, 2004). Furthermore, multiple copies of these genes might occur within a single organism (van Beilen & Funhoff, 2007). It has been proposed that the clustering of AlkB sequences in a phylogenetic tree might be based on their substrate specificity rather than the phylogenetic origin of bacterial groups carrying these genes (Heiss-Blanquet et al, 2005; Van Beilen et al, 2003).

It is noteworthy that most of the members of early phase AlkB communities belonged to the same cluster in the phylogenetic tree (Fig 8, **II**) indicating phylogenetic closeness and similar substrate specificity of these groups. These early phase members showed a high sequence similarity to *Rhodococcus sp*. MS103a (83 – 91% amino acid sequence similarity) and also to *Pedobacter sp.* MS245e (85%). These observations are congruent with the dominance of Actinomycetales in the early phase 16S rRNA gene communities and suggest the potential role of Actinomycetales as alkane degraders in the early stages of oil pollution. In the late phase however, AlkB communities were more diverse as they spanned across 9 different clusters in the phylogenetic tree. These findings probably point towards the increased diversity related to substrate specificity for alkanes in the later stages of succession.

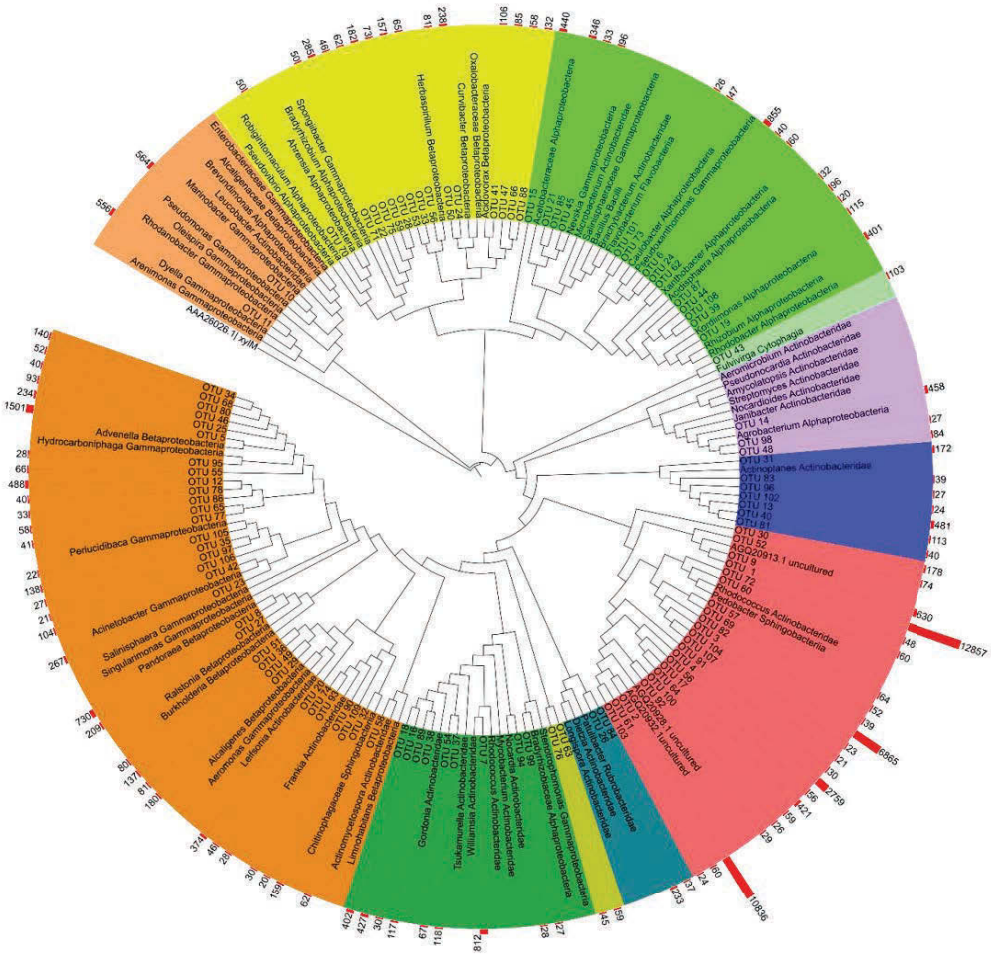


Fig. 15: Maximum likelihood tree of alkane monooxygenase (AlkB) amino-acid sequences based on the JTT model (Jones et al, 1992). The bootstrap consensus tree was inferred from 500 replicates. One representative sequence from each AlkB OTUs was included in the tree along with the closest hits from NCBI non-redundant protein database. OTUs with more than 20 sequences were chosen for better visualization. The red bars indicate the abundance of OTUs (number of sequences).

4.2. Spatial patterns in the creosote-contaminated site (III)

We studied an aged creosote-contaminated site in order to elucidate the spatial patterns of microbial diversity and activity and to relate them with the soil-chemical parameters on the site.

4.2.1. Niche differentiation explaining the spatial patterns of bacterial diversity

We used an intensive grid based sampling scheme followed by 454 pyrosequencing of 16S rRNA gene amplicons and geostatistical modelling in order to relate the spatial patterns of soil

chemical parameters and bacterial diversity at the studied creosote-polluted site. Kriged maps revealed two distinct patterns in the soil chemistry of the site (Fig. 1**, III**): a) patchy distribution of pollutants (TPHs and PAHs) and b) east to west pH gradient (7.8 - 5). Also, bacterial community structure in the polluted site was mainly regulated by the level of pollution and pH (Fig. 2, **III**). Bacterial diversity was found to be significantly lower in the hotspots of pollution (Spearman's r= -0.45, p= 0.008).

We further tried to analyse whether the spatial distribution of bacterial taxa could be related to the spatial variation of soil properties. Semivariogram modelling and mapping of relative abundance of bacterial taxa provided insights into the patterns of niche partitioning at different taxonomic levels regulated by the concentration of creosotes and pH levels. Proteobacteria displayed a patchy distribution pattern and dominated in the hotspots of pollution. Patchiness in spatial distribution was also displayed by Actinobacteria, TM7, Armatimonadetes and Planctomycetes but the relative abundances of these groups were significantly reduced in the hotspots of pollution (Fig 3, **III**). On the other hand, spatial patterns of Acidobacteria, Verrucomicrobia and Bacteriodetes were found to be related to the pH gradient. A considerable dominance of Proteobacteria over other phyla in the hotspots of pollution is supported by the fact that this phylum consists of large number of well-known hydrocarbon degraders. However, the drastic decrease in actinobacterial abundance in the hotspots of pollution contradicts our knowledge of the robust catabolic capacities of actinobacterial groups (Kanaly & Harayama, 2000; Larkin et al, 2005; Arenskötter et al, 2004) and some previous findings that have reported the dominance of Actinobacteria in hydrocarbon polluted soil (Antizar-Ladislao et al, 2008; Johnsen et al, 2007). Interestingly, a similar study on an aged creosote polluted site in Sweden has also reported a decline in actinobacterial abundance in the hotspots of pollution (Törneman et al, 2008). These findings indicate that Actinobacteria were outcompeted by Proteobacteria over time in this polluted site.

We also looked into bacterial responses to environmental factors at lower taxonomic levels (Fig 4, **III**). Among the proteobacterial classes, Betaproteobacteria and Gammaproteobacteria had increased relative abundances in the hotspots of pollution while Deltaproteobacteria showed an opposite trend. Spatial distribution of Alphaproteobacteria could not be explained by any of the measured soil chemical parameters. The tremendous diversity of niche specialization patterns within alphaproteobacterial orders is well known (Ettema & Andersson, 2009). Alphaproteobacterial orders, Sphingomonadales and Rhodospirillales were more abundant in high pollution level, whereas, Rhizobiales followed the opposite trend. pH also played an important role in regulating the spatial patterns of proteobacterial groups dominating in high pollution. Sphingomonadales prevailed in high pH whereas Rhodospirillales preferred lower pH zones in the polluted site. Similarly, betaproteobacterial family Burkholderiaceae was found in higher abundances in low pH and Pseudomonadaeceae (Gammaproteobacteria) were prevalent in higher pH. Several genera of Rhodospiralles have previously been linked with hydrocarbon degradation at moderately acidic pH levels (Röling et al, 2006; Hamamura et al, 2005). Although, Burkholderiaceae and Pseudomonadaceae are among some of the most well-known hydrocarbon degraders in a variety of soils (Leys et al, 2004; Pérez-Pantoja et al, 2012; Silby et al, 2011; Stolz, 2013), the pH optima of these groups is in polluted soils is still unaccounted for. Our results provide some clues about the pH-specific niche preferences of these groups in aged hydrocarbon contamination. Acidobacterial classes displayed contrasting patterns of pH optima in the studied site. Acidobacteria Gp1, Gp2 and Gp3 were favoured in low pH, whereas Acidobacteria

Gp 4 and Gp 6 prevailed in higher pH. These observations are in line with the findings from some recent studies (Foesel et al, 2013; George et al, 2011; Rousk et al, 2010). Acidobacteria came across as generalists in hydrocarbon polluted soil as they appeared to remain unaffected by the varying pollution levels. It is worth mentioning here that role of Acidobacteria as generalist group in pollution was also highlighted in the other two studies (**I&II**). Combining the results of spatial patterns of bacterial diversity with soil chemistry not only provided a thorough knowledge of long term effects of creosote pollution but also brought forward the possibility of utilizing the spatial patterns displayed by specific bacterial groups as proxies in monitoring polluted sites.

4.2.2. Spatial patterns of microbial activity

Assessment of microbial activity in the polluted site was performed by basal respiration assay and fluorescein diacetate (FDA) hydrolysis assay. Basal respiration assay has been found to be a relevant measure of total biological activity in soils and has previously been used as a parameter for monitoring hydrocarbon bioremediation (Margesin et al, 2003; Tuomi et al, 2004). FDA hydrolysis provides a general measure of soil hydrolytic activity as it measures the activity of a broad range of enzymes (proteases, lipases and esterases) capable of cleaving the fluorogenic FDA substrate into fluorescein (Schnurer & Rosswall, 1982); this assay has also been used in hydrocarbon bioremediation studies (Song & Bartha, 1990; Wang & Bartha, 1990; Margesin et al, 2003). The basal respiration rate in the polluted site varied from 0.1 to 1.1 µl CO_2/g soil/ hour and the FDA hydrolysis rate was in the range of 19.6 to 140.5 µg fluorescein/ g soil 3h (Fig 5, **III**). Both measures of microbial activity showed a high correlation with each other ($r=0.75$, $p<0.0001$) and were positively correlated to the concentration of different PAHs and TPHs (Supplementary Table S3, **III**). Higher microbial activity and reduced diversity in the most polluted zones in this aged contaminated site could be an indication of adaptation or enrichment of specific indigenous microbes such as proteobacterial groups dominating in pollution and successful utilization of hydrocarbons as a food source by these groups.

5. Conclusions

Secondary succession and spatial patterns of microbes in hydrocarbon contaminated ecosystems (soils and *Populus* rhizosphere) were the two main topics explored in this thesis. Poplars have been extensively used for rhizoremediation of organic contaminants but surprisingly few studies have focused on its rhizosphere-associated microbial communities in polluted soils.

It was evident from both studies on secondary succession that oil pollution initiated deterministic patterns of temporal changes in soil and rhizosphere-associated bacterial communities. Structural and functional bacterial communities were resilient to oil pollution as shown by the steady rise in diversity after the initial deleterious effect of hydrocarbon components. In the long-term study of secondary succession, distinct changes occurred in the composition of structural (16S rRNA) and catabolic (extradiol dioxygenases and alkane monooxygenases) communities in the early and late phases. A transition from low diversity-high dominance community to high diversity-low dominance groups occurred in the late phase of succession. *Sphingomonas* type extradiol dioxygenases and *Rhodococcus* type alkane monooxygenases dominated the early phase of succession whereas the late phase members of both the catabolic communities were phylogenetically diverse.

Among the two catabolic genes studied, alkane monooxygenase databases were far more developed compared to those of extradiol dioxygenases. The identification of putative hosts of *alkB* sequences was, however, not so straight forward due to the incongruence of *alkB* and 16S phylogenies, most probably caused due to the extensive horizontal gene transfer. It was hard to draw conclusions about the congruence of 16S rRNA and extradiol dioxygenases based phylogeny as very few clusters found in our study contained cultured representatives. This underpins the need for much extensive isolation of aromatic degraders and characterization of enzymes involved in the upper *meta*-pathway of aromatic degradation.

Niche differentiation based on creosote concentration and pH was an important mechanism explaining the spatial patterns of bacterial taxa in the aged creosote contaminated site. Proteobacterial groups were the key groups dominating in high levels of creosote contamination. To our surprise, Actinobacteria, a phyla which consists of many robust hydrocarbon degraders, dramatically decreased in abundance in high creosote pollution. Interestingly, in our 2-year field study of bacterial succession, abundance of actinobacterial groups was also considerably reduced in the late phase of succession. Taken together, these findings lead us to suggest that actinobacterial groups might be the r-strategists in hydrocarbon polluted soils and are less important in the later stages of secondary succession. According to the results obtained in this thesis, Acidobacteria represent the generalists in hydrocarbon contaminated soils.

Geostatistical approaches hold a great promise for characterizing the microbial parameters on polluted sites. Integrating microbial and chemical analyses in the characterization of an aged polluted site could provide a more comprehensive site pre-evaluation for bioremediation. Prospects of using spatial patterns of specific bacterial taxa as proxies in monitoring polluted sites should be further explored. It is highly relevant to target specific catabolic genes along with the phylogenetic markers in order to gain a detailed understanding of microbial community dynamics in polluted ecosystems. It is noteworthy that

the level of resolution achieved in profiling of structural and catabolic microbial communities in this study could be achieved mostly due to the advancements in next generation sequencing technology.

6. Acknowledgements

"Clearly, one of the major obstacles to our experience of gratitude is the habit we have of sleepwalking through life. The truth is that we are never lacking for blessings in our lives, but we are often lacking in awareness and recognition of them." ~ Dr. Diane Berke

It is strange, this realization, that I am writing the last chapter of my PhD thesis. *Strange*, because I have been in relentless pursuit of this very moment for the past four years but right when this moment is upon me, I am struggling to find words; words, that could embody the depth of the countless contributions that have made this endeavour possible.

This thesis work was carried out at the Division of General Microbiology, Department of Biosciences at the University of Helsinki and was financially supported by Maj and Tor Nessling Foundation and University of Helsinki Research Foundation. I take this opportunity to acknowledge the funding bodies, and the excellent research facilities and staff at the Division of General Microbiology that played an instrumental role in successful completion of this work.

I wish to thank my supervisor, Dr. Kim Yrjälä, for his able mentorship during my PhD. Kim, I am grateful for your constant support and enthusiastic guidance. The endless opportunities of networking and research exchanges you have provided to me during these years, have definitely enabled me to broaden my horizons and to develop myself as a better researcher.

I sincerely thank the members of my thesis advisory committee- Dr. Kirsten Jørgensen and Dr. Merja Itävaara for their regular and useful feedback on my thesis work. Also, I would like to thank Professor Jaak Truu and Professor Tomas Macek for kindly reviewing this thesis.

I wish to express my gratitude towards Dr. Pertti Pulkkinen for providing excellent facilities for carrying out the rhizoremediation experiments, and for his expert guidance on statistics and experimental setup. Raimo, I thank you for the expertise you lent in setting up the greenhouse and field study, for arranging the sampling trips and for being patient with my never ending technical questions. I would like to thank the staff of METLA who took care of aspen plants during the field experiment and the trainees who helped me with sampling (specially the brave ones who were freezing with me during winter sampling!). The team from Golder associates and METLA are thanked for carrying out the extensive sampling at the Luumäki site. I would also like to acknowledge Dr. Pirjo Tuomi for providing access to the geochemical data of Luumäki site and also for sharing her knowledge on the site history.

I would also like to thank the DNA Sequencing and Genomics lab at the University of Helsinki for the successful completion of 4 pyrosequencing projects for this thesis. I would like to thank, Lars Paulin for his expert guidance in library preparation. Kaisa, thank you for answering all my questions, both sequencing and thesis related.

I thank the graduate schools- EnSTe and MBDP for organizing useful courses and for all the travel grants they awarded to me, which made my doctoral studies a great learning and networking experience. I was fortunate to be a part of scientific networks of NENUN and EU-COST; the workshops and research exchange programs within these networks have contributed immensely towards improving the quality of my thesis.

I would like to take this opportunity to thank the previous and present members of MEM group- Mirja, Timo, Heli, Anuliina, Sajan, Magali, Amelie, Faiza, Cosme, Eugenia, Alice, Hyunseok, Pauli,

Alex, Ludwig and Nelson for making the work environment so motivational. Mirja, you have been on my side from the beginning of my PhD; from initially training me in the lab to this date, planning my Karonkka, and I am deeply thankful to you. Timo, you were my "trouble-shooter"; may it be a tricky PCR or complicated data analysis, you always had a solution for everything!☺ I would like to thank you, Heli, for being almost a second supervisor for me. You were patient enough to go through all my texts, correcting the same mistakes that I kept repeating, helping me to turn the "lost in correlations" manuscript into an article with "innovative statistical approaches"; you have always been of great help. I must admit, at some point, I had started to use you as an alternate "Google search engine", shooting every question that came to my mind at you, right away!☺ Apart from the fun at work, it was a great pleasure to have those awesome game-evenings with Mirja, Heli, Anu and Timo. I hope that at least some of our highly ambitious "business plans" will work someday, I am tempted to name a few here but I will hold back. ☺

Being far away from home was stressful at times but I was lucky enough to have met some wonderful people in Helsinki who made this journey enjoyable. Poojitha and Tuomas- thank you guys for being a "surrogate" family for me, and keeping "my den" ready whenever I came to stay over.☺ Aura- I would like to thank you for the constant support since the day I moved to Finland, not to forget, you were the one who told me about this PhD position.☺ We have had some great time while taking long walks in the autumn, discussing the passion for cooking that we share, sledging, film festivals and the "academic tourism" in Czech Republic, Iceland and Denmark. Sofia- I thank you for the awesome time that we have had together, both on and off campus. Ajay- apart from being my MATLAB and R guru, you also were a great company for sports and for sci-fi, political and nerdy discussions which often shaped into incessant debates.☺ I wish to thank, Jaana and Kim, for the lovely dinners at Christmas, Easter and many other festive occasions every year and the fantastic trip to Bromarv.☺ Sorina, Yoann, Naveen, Zohaib, Efi, Grit and Windi- thank you guys, for all the fun get-togethers that made my time in Helsinki so vivid.

Finally, I wish to express my heartfelt gratitude towards my family. Maa and Baba- I deeply thank you for believing in my dreams, for respecting my life choices, for your sacrifices, for giving me all the independence and at the same time, the much needed emotional strength through this journey. Neal- you are not only my younger brother but also my best friend; thank you for cheering me up during my low moments.☺ I thank my grandmother and late grandfather, for inculcating in me, qualities of perseverance and discipline, which played a big role in this achievement. My uncles Raja and Amit and my aunt Reeta are thanked for their enthusiastic support and motivation during all these years. I would also like to thank my parent in-laws and Dhananjay, for their constant encouragement and support. My warmest thanks go to my husband Abhishek for his endless emotional support, care and love. Being a skilled researcher himself, he has taught me many "tricks of the trade", inspired me to raise the bar and stood by me through thick and thin. I do not think I will ever be able to find enough words for thanking you, but you do know that you mean the world to me!

10th August, 2014

7. References

Abdo Z, Schüette UM, Bent SJ, Williams CJ, Forney LJ, Joyce P (2006). Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. Environ Microbiol 8:929-938.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533-538.

Andreolli M, Lampis S, Poli M, Gullner G, Biró B, Vallini G (2013). Endophytic *Burkholderia fungorum* DBT1 can improve phytoremediation efficiency of polycyclic aromatic hydrocarbons. Chemosphere 92:688-694.

Andria V, Reichenauer TG, Sessitsch A (2009). Expression of alkane monooxygenase *alk*B genes by plant-associated bacteria in the rhizosphere and endosphere of Italian ryegrass *Lolium multiflorum* L. grown in diesel contaminated soil. Environ pollut 157:3347-3350.

Andujar E, Hernaez MJ, Kaschabek SR, Reineke W, Santero E (2000). Identification of an extradiol dioxygenase involved in tetralin biodegradation: gene sequence analysis and purification and characterization of the gene product. J Bacteriol 182:789-795.

Antizar-Ladislao B, Spanova K, Beck AJ, Russell NJ (2008). Microbial community structure changes during bioremediation of PAHs in an aged coal-tar contaminated soil by in-vessel composting. Int Biodeterior Biodegrad 61:357-364.

Arenskötter M, Bröker D, Steinbüchel A (2004). Biology of the metabolically diverse genus Gordonia. Appl Environ Microbiol 70:3195-3204.

Arnold C, Parfitt D, Kaltreider M (2007). Phytovolatilization of oxygenated gasoline-impacted groundwater at an underground storage tank site via conifers. Int J Phytoremediation 9:53-69.

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006). The role of root exudates in rhizosphere interactions with plants and other organisms. Annu.Rev.Plant Biol. 57:233-266.

Bamforth SM, Singleton I (2005). Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. J Chem Technol and Biot 80:723-736.

Bannert A, Kleineidam K, Wissing L, Mueller-Niggemann C, Vogelsang V, Welzl G*, et al.* (2011). Changes in diversity and functional gene abundances of microbial communities involved in nitrogen fixation, nitrification, and denitrification in a tidal wetland versus paddy soils cultivated for different time periods. Appl Environ Microbiol 77:6109-6116.

Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL*, et al.* (2011). Soil Microbial Community Successional Patterns during Forest Ecosystem Restoration. Appl Environ Microbiol 77:6158-6164.

Barberán A, Bates ST, Casamayor EO, Fierer N (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6:343-351.

Bell SG, Dale A, Rees NH, Wong L (2010). A cytochrome P450 class I electron transfer system from Novosphingobium aromaticivorans. Appl Microbiol Biotechnol 86:163-175.

Bell SG, Hoskins N, Xu F, Caprotti D, Rao Z, Wong L (2006). Cytochrome P450 enzymes from the metabolically diverse bacterium< i> Rhodopseudomonas palustris. Biochem Biophys Res Commun 342:191-196.

Bell TH, Yergeau E, Martineau C, Juck D, Whyte LG, Greer CW (2011). Identification of Nitrogen-Incorporating Bacteria in Petroleum-Contaminated Arctic Soils by Using [15N] DNA-Based Stable Isotope Probing and Pyrosequencing. Appl Environ Microbiol 77:4163-4171.

Bengtsson G, Törneman N, Yang X (2010). Spatial uncoupling of biodegradation, soil respiration, and PAH concentration in a creosote contaminated soil. Environ pollut 158:2865-2871.

Bispo A, Jourdain MJ, Jauzein M (1999). Toxicity and genotoxicity of industrial soils polluted by polycyclic aromatic hydrocarbons (PAHs). Org Geochem 30:947-952.

Boonsaner M, Borrirukwisitsak S, Boonsaner A (2011). Phytoremediation of BTEX contaminated soil by *Canna×generalis*. Ecotoxicol Environ Saf 74:1700-1707.

Boopathy R (2000). Factors limiting bioremediation technologies. Bioresour Technol 74:63-67.

Bradshaw H, Ceulemans R, Davis J, Stettler R (2000). Emerging model systems in plant biology: poplar (Populus) as a model forest tree. J Plant Growth Regul 19:306-313.

Brockman FJ, Murray CJ (1997). Subsurface microbiological heterogeneity: current knowledge, descriptive approaches and applications. FEMS Microbiol Rev 20:231-247.

Bru D, Ramette A, Saby N, Dequiedt S, Ranjard L, Jolivet C*, et al.* (2010). Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. ISME J 5:532-542.

Burken JG, Ross C, Harrison LM, Marsh A, Zetterstrom L, Gibbons JS (2001). Benzene toxicity and removal in laboratory phytoremediation studies. Pract Periodical Hazard, Toxic, Radioact Waste Manage 5:161-171.

Burken JG, Schnoor JL (1996). Phytoremediation: plant uptake of atrazine and role of root exudates. J Environ Eng 122:958-963.

Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, Kjelleberg S (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73:278-288.

Cerniglia CE (1993). Biodegradation of polycyclic aromatic hydrocarbons. Curr Opin Biotechnol 4:331-338.

Chapin III FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL*, et al.* (2000). Consequences of changing biodiversity. Nature 405:234-242.

Connell JH, Slatyer RO (1977). Mechanisms of succession in natural communities and their role in community stability and organization. Am Nat: 1119-1144.

Cook RL, Hesterberg D (2013). Comparison of trees and grasses for rhizoremediation of petroleum hydrocarbons. Int J Phytoremediation 15:844-860.

Črešnar B, Petrič Š (2011). Cytochrome P450 enzymes in the fungal kingdom. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1814:29-35.

Crosby LD, Criddle CS (2003). Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. BioTechniques 34:790-803.

Cunningham SD, Berti WR, Huang JW (1995). Phytoremediation of contaminated soils. Trends Biotechnol 13:393-397.

Curtis TP, Sloan WT (2005). Microbiology. Exploring microbial diversity--a vast below. Science 309:1331-1333.

De Carcer DA, Martin M, Karlson U, Rivilla R (2007). Changes in bacterial populations and in biphenyl dioxygenase gene diversity in a polychlorinated biphenyl-polluted soil after introduction of willow trees for rhizoremediation. Appl Environ Microbiol 73:6224.

Dec J, Bollag J (1994). Use of plant material for the decontamination of water polluted with phenols. Biotechnol Bioeng 44:1132-1139.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM*, et al.* (2008). Functional metagenomic profiling of nine biomes. Nature 452:629-632.

Dohrmann A, Tebbe C, Kowalchuk G, de Bruijn F, Head I, Akkermans A*, et al.* (2004). Microbial community analysis by PCR-single-strand conformation polymorphism (PCR-SSCP). Molecular microbial ecology manual. Volumes 1 and 2:809-838.

Dollhopf S, Hashsham S, Tiedje J (2001). Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal component analysis to describe community dynamics and convergence. Microb Ecol 42:495-505.

Donachie SP, Foster JS, Brown MV (2007). Culture clash: challenging the dogma of microbial diversity. ISME J 1:97-99.

Dong W, Eichhorn P, Radajewski S, Schleheck D, Denger K, Knepper TP*, et al.* (2004). Parvibaculum lavamentivorans converts linear alkylbenzenesulphonate surfactant to sulphophenylcarboxylates, α, β-unsaturated sulphophenylcarboxylates and sulphophenyldicarboxylates, which are degraded in communities. J Appl Microbiol 96:630-640.

Doni S, Macci C, Peruzzi E, Arenella M, Ceccanti B, Masciandaro G (2012). In situ phytoremediation of a soil historically contaminated by metals, hydrocarbons and polychlorobiphenyls. J Environ Monitor 14:1383-1390.

Dos Santos HF, Cury JC, Do Carmo FL, Dos Santos AL, Tiedje J, van Elsas JD*, et al.* (2011). Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. PLoS One 6:e16943.

Douben PE (2003). Introduction. In: PAHs: An Ecotoxicological Perspective. John Wiley & Sons, Ltd, pp 1-6.

Doucette W, Bugbee B, Smith S, Pajak C, Ginn J (2003). Uptake, metabolism, and phytovolatilization of trichloroethylene by indigenous vegetation: Impact of precipitation. Phytoremediation: Transformation and Control of Contaminants, Wiley/Interscience, Hoboken, NJ: 561-588.

Dubinsky EA, Conrad ME, Chakraborty R, Bill M, Borglin SE, Hollibaugh JT*, et al.* (2013). Succession of Hydrocarbon-Degrading Bacteria in the Aftermath of the Deepwater Horizon Oil Spill in the Gulf of Mexico. Environ Sci Technol 47:10860-10867.

Dunfield KE, Germida JJ (2003). Seasonal Changes in the Rhizosphere Microbial Communities Associated with Field-Grown Genetically Modified Canola (Brassica napus). Appl Environ Microbiol 69:7310-7318.

Edwards R, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson D*, et al.* (2006). Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7:57.

Enwall K, Throbäck IN, Stenberg M, Söderström M, Hallin S (2010). Soil resources influence spatial patterns of denitrifying communities at scales compatible with land management. Appl Environ Microbiol 76:2243-2250.

Ettema CH, Wardle DA (2002). Spatial soil ecology. Trends in ecology & evolution 17:177-183.

Ettema TJG, Andersson SGE (2009). The a-proteobacteria: the Darwin finches of the bacterial world. Biol Lett 5:429-432..

Feng L, Wang W, Cheng J, Ren Y, Zhao G, Gao C*, et al.* (2007). Genome and proteome of long-chain alkane degrading Geobacillus thermodenitrificans NG80-2 isolated from a deep-subsurface oil reservoir. Proc Natl Acad Sci U S A 104:5602-5607.

Fierer N, Nemergut D, Knight R, Craine JM (2010). Changes through time: integrating microorganisms into the study of succession. Res Microbiol 161:635-642.

Fisher MM, Triplett EW (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl Environ Microbiol 65:4630-4636.

Fletcher JS, Hegde RS (1995). Release of phenols by perennial plant roots and their potential importance in bioremediation. Chemosphere 31:3009-3016.

Foesel BU, Nägele V, Naether A, Wüst PK, Weinert J, Bonkowski M*, et al.* (2014). Determinants of Acidobacteria activity inferred from the relative abundances of 16S rRNA transcripts in German grassland and forest soils. Environ Microbiol 16:658-675.

Fox BG, Borneman JG, Wackett LP, Lipscomb JD (1990). Haloalkene oxidation by the soluble methane monooxygenase from Methylosinus trichosporium OB3b: mechanistic and environmental implications. Biochem (N Y) 29:6419-6427.

Fuchs G, Boll M, Heider J (2011). Microbial degradation of aromatic compounds—from one strategy to four. Nat Rev Microbiol 9:803-816.

Fuhrman JA (2009). Microbial community structure and its functional implications. Nature 459:193-199.

Funhoff EG, Bauer U, Garcia-Rubio I, Witholt B, van Beilen JB (2006). CYP153A6, a soluble P450 oxygenase catalyzing terminal-alkane hydroxylation. J Bacteriol 188:5220-5227.

George IF, Hartmann M, Liles MR, Agathos SN (2011). Recovery of as-yet-uncultured soil Acidobacteria on dilute solid media. Appl Environ Microbiol 77:8184-8188.

Gerhardt KE, Huang XD, Glick BR, Greenberg BM (2009). Phytoremediation and rhizoremediation of organic soil contaminants: Potential and challenges. Plant science 176:20-30.

Gibson J, S. Harwood C (2002). Metabolic diversity in aromatic compound utilization by anaerobic microbes. Annual Reviews in Microbiology 56:345-369.

Giebler J, Wick LY, Chatzinotas A, Harms H (2013a). Alkane-degrading bacteria at the soil–litter interface: comparing isolates with T-RFLP-based community profiles. FEMS Microbiol Ecol 86:45-58.

Giebler J, Wick LY, Schloter M, Harms H, Chatzinotas A (2013b). Evaluating the Assignment of alkB Terminal Restriction Fragments and Sequence Types to Distinct Bacterial Taxa. Appl Environ Microbiol 79:3129-3132.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA (2010). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME J 5:461-472.

Goodison S, Urquidi V, Kumar D, Reyes L, Rosser CJ (2013). Complete genome sequence of Mycoplasma hyorhinis strain SK76. Genome announcements 1.

Goovaerts P (1999). Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89:1-45.

Goovaerts P (1998). Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biol Fertility Soils 27:315-334.

Goovaerts P (2001). Geostatistical modelling of uncertainty in soil science. Geoderma 103:3-26

Gordon M, Choe N, Duffy J, Ekuan G, Heilman P, Muiznieks I*, et al.* (1998). Phytoremediation of trichloroethylene with hybrid poplars. Environ Health Perspect 106 Suppl 4:1001-1004.

Gotelli NJ, McCabe DJ (2002). Species co-occurrence: a meta-analysis of JM Diamond's assembly rules model. Ecology 83:2091-2096.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M*, et al.* (2004). Spatial scaling of microbial eukaryote diversity. Nature 432:747-750.

Griffiths BS, Philippot L (2012). Insights into the resistance and resilience of the soil microbial community. FEMS Microbiol Rev. 37:112-129.

Hamamura N, Olson SH, Ward DM, Inskeep WP (2005). Diversity and Functional Analysis of Bacterial Communities Associated with Natural Hydrocarbon Seeps in Acidic Soils at Rainbow Springs, Yellowstone National Park. Appl Environ Microbio 71:5943-5950.

Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews 68:669-685.

Harms H, Schlosser D, Wick LY (2011). Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. Nat Rev Microbiol 9:177-192.

Harvey RG (1996). Mechanisms of carcinogenesis of polycyclic aromatic hydrocarbons. Polycyclic Aromatic Compounds 9:1-23.

Head IM, Jones DM, Röling WF (2006). Marine microorganisms make a meal of oil. Nat Rev Microbiol 4:173-182.

Heider J, Fuchs G (1997). Anaerobic Metabolism of Aromatic Compounds. European Journal of Biochemistry 243:577-596.

Heiss-Blanquet S, Benoit Y, Marechaux C, Monot F (2005). Assessing the role of alkane hydroxylase genotypes in environmental samples by competitive PCR. J Appl Microbiol 99:1392-1403.

Hirsch PR, Mauchline TH, Clark IM (2010). Culture-independent molecular techniques for soil microbial ecology. Soil Biol Biochem 42:878-887.

Hirsh S, Compton H, Matey D, Wrobel J, Schneider W (2003). Five-Year Pilot Study: Aberdeen Proving Ground, Maryland. Phytoremediation: Transformation and control of Contaminants: 635-659.

Hong MS, Farmayan WF, Dortch IJ, Chiang CY, McMillan SK, Schnoor JL (2001). Phytoremediation of MTBE from a groundwater plume. Environ Sci Technol 35:1231-1239.

Hong SH, Ryu H, Kim J, Cho K (2011). Rhizoremediation of diesel-contaminated soil using the plant growth-promoting rhizobacterium Gordonia sp. S2RP-17. Biodegradation 22:593-601.

Horn HS (1974). The ecology of secondary succession. Annu Rev Ecol Syst 5:25-37.

Horner-Devine MC, Silver JM, Leibold MA, Bohannan BJ, Colwell RK, Fuhrman JA*, et al.* (2007). A comparison of taxon co-occurrence patterns for macro-and microorganisms. Ecology 88:1345-1353.

Hugenholtz P (2002). Exploring prokaryotic diversity in the genomic era. Genome Biol 3:1-0003.8.

Iida T, Sumita T, Ohta A, Takagi M (2000). The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, Yarrowia lipolytica: cloning and characterization of genes coding for new CYP52 family members. Yeast 16:1077-1087.

Jennings AA, Petersen EJ (2006). Variability of North American regulatory guidance for heavy metal contamination of residential soil. J Environ Eng Sci 5:485-508.

Jeon CO, Park M, Ro HS, Park W, Madsen EL (2006). The naphthalene catabolic (nag) genes of Polaromonas naphthalenivorans CJ2: evolutionary implications for two gene clusters and novel regulatory control. Appl Environ Microbiol 72:1086-1095.

Johnsen AR, Schmidt S, Hybholt TK, Henriksen S, Jacobsen CS, Andersen O (2007). Strong Impact on the Polycyclic Aromatic Hydrocarbon (PAH)-Degrading Community of a PAH-Polluted Soil but Marginal Effect on PAH Degradation when Priming with Bioremediated Soil Dominated by Mycobacteria. Appl Environ Microbiol 73:1474-1480.

Jones DT, Taylor WR, Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS 8:275-282.

Jordahl JL, Foster L, Schnoor JL, Alvarez PJ (1997). Effect of hybrid poplar trees on microbial populations important to hazardous waste bioremediation. Environ Toxicol Chem 16:1318-1321.

Jouanneau Y, Meyer C (2006). Purification and characterization of an arene cis-dihydrodiol dehydrogenase endowed with broad substrate specificity toward polycyclic aromatic hydrocarbon dihydrodiols. Appl Environ Microbiol 72:4726-4734.

Juottonen H, Hynninen A, Nieminen M, Tuomivirta TT, Tuittila E, Nousiainen H*, et al.* (2012). Methane-cycling microbial communities and methane emission in natural and restored peatlands. Appl Environ Microbiol 78:6386-6389.

Jurelevicius D, Alvarez VM, Peixoto R, Rosado AS, Seldin L (2013). The use of a combination of alkB primers to better characterize the distribution of alkane-degrading bacteria. PloS one 8:e66565.

Kanaly RA, Harayama S (2000). Biodegradation of High-Molecular-Weight Polycyclic Aromatic Hydrocarbons by Bacteria. J Bact 182:2059-2067.

Kaplan CW, Kitts CL (2004). Bacterial succession in a petroleum land treatment unit. Appl Environ Microbiol 70:1777-1786.

Keck A, Conradt D, Mahler A, Stolz A, Mattes R, Klein J (2006). Identification and functional analysis of the genes for naphthalenesulfonate catabolism by Sphingomonas xenophaga BN6. Microbiol 152:1929-1940.

Kloos K, Munch JC, Schloter M (2006). A new method for the detection of alkane-monooxygenase homologous genes *alkB* in soils based on PCR-hybridization. J Microbiol Methods 66:486-496.

Klopfenstein NB, YoungWoo C, MeeSook K, Ahuja MR, Dillon MC, Carman RC*, et al.* (1997). Micropropagation, genetic engineering, and molecular biology of Populus. In: General Technical Report RM-GTR-297, p. 326, US Department of Agriculture, Forest Service, Rocky Mountain Research Station

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R*, et al.* (2011). Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A 108 Suppl 1:4578-4585.

Kotik M, Davidová A, Voříšková J, Baldrian P (2013). Bacterial communities in tetrachloroethene-polluted groundwaters: A case study. Sci Total Environ 454:517-527.

Kowalchuk GA, Stienstra AW, Heilig, Stephen JR, Woldendorp JW (2000). Changes in the community structure of ammonia-oxidizing bacteria during secondary succession of calcareous grasslands. Environ Microbiol 2:99-110.

Krause S, Lüke C, Frenzel P (2009). Spatial heterogeneity of methanotrophs: a geostatistical analysis of pmoA-based T-RFLP patterns in a paddy soil. Environmental Microbiology Reports 1:393-397.

Krige D (1976). A review of the development of geostatistics in South Africa. In: Advanced Geostatistics in the Mining Industry. Springer, pp 279-293.

Kuiper I, Lagendijk EL, Bloemberg GV, Lugtenberg BJJ (2004). Rhizoremediation: a beneficial plant-microbe interaction. Mol Plant-Microbe Interact 17:6-15.

Larkin MJ, Kulakov LA, Allen CC (2005). Biodegradation and Rhodococcus–masters of catabolic versatility. Curr Opin Biotechnol 16:282-290.

Lauber CL, Hamady M, Knight R, Fierer N (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75:5111-5120.

Leahy JG, Colwell RR (1990). Microbial degradation of hydrocarbons in the environment. Microbiological Reviews 54:305-315.

Lee KY, Strand SE, Doty SL (2012). Phytoremediation of chlorpyrifos by populus and salix. Int J Phytoremediation 14:48-61.

Lehembre F, Doillon D, David E, Perrotto S, Baude J, Foulon J*, et al.* (2013). Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. Environ Microbiol 15:2829-2840.

Leigh MB, Pellizari VH, Uhlík O, Sutka R, Rodrigues J, Ostrom NE*, et al.* (2007). Biphenyl-utilizing bacteria and their functional genes in a pine root zone contaminated with polychlorinated biphenyls (PCBs). ISME J 1:134-148.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol G*, et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806-809.

Leys NM, Ryngaert A, Bastiaens L, Verstraete W, Top EM, Springael D (2004). Occurrence and phylogenetic diversity of Sphingomonas strains in soils contaminated with polycyclic aromatic hydrocarbons. Appl Environ Microbiol 70:1944-1955.

Li L, Liu X, Yang W, Xu F, Wang W, Feng L*, et al.* (2008). Crystal structure of long-chain alkane monooxygenase (LadA) in complex with coenzyme FMN: unveiling the long-chain alkane hydroxylase. J Mol Biol 376:453-465.

Lipscomb JD (2008). Mechanism of extradiol aromatic ring-cleaving dioxygenases. Curr Opin Struct Biol 18:644-649.

Lipscomb JD (1994). Biochemistry of the soluble methane monooxygenase. Annual Reviews in Microbiology 48:371-399.

Liu J, Schnoor JL (2008). Uptake and translocation of lesser-chlorinated polychlorinated biphenyls (PCBs) in whole hybrid poplar plants after hydroponic exposure. Chemosphere 73:1608-1616.

Liu WT, Marsh TL, Cheng H, Forney LJ (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl Environ Microbio 63:4516-4522.

Lu J, Nomura N, Nakajima-Kambe T, Nakahara T (2000). Cloning and expression of genes encoding< i> meta-cleavage enzymes from 4, 6-dimethyldibenzothiophene-degrading *Sphingomonas* strain TZS-7. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1492:493-498.

Macek T, Mackova M, Káš J (2000). Exploitation of plants for the removal of organics in environmental remediation. Biotechnol Adv 18:23-34.

Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ*, et al.* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480:368-371.

Mackova M, Prouzova P, Stursa P, Ryslava E, Uhlik O, Beranova K*, et al.* (2009). Phyto/rhizoremediation studies using long-term PCB-contaminated soil. Environ Sci Poll Res 16:817-829.

Maeng JH, Sakai Y, Ishige T, Tani Y, Kato N (1996). Diversity of dioxygenases that catalyze the first step of oxidation of long-chain n-alkanes in Acinetobacter sp. M-1. FEMS Microbiol Lett 141:177-182.

Margesin R, Labbe D, Schinner F, Greer C, Whyte L (2003). Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine alpine soils. Appl Environ Microbiol 69:3085-3092.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA*, et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Marshall AG, Rodgers RP (2004). Petroleomics: The next grand challenge for chemical analysis. Acc Chem Res 37:53-59.

Martin BC, George SJ, Price CA, Ryan MH, Tibbett M (2014). The role of root exuded low molecular weight organic anions in facilitating petroleum hydrocarbon degradation: Current knowledge and future directions. Sci Total Environ 472:642-653.

McDonald IR, Murrell JC (1997). The particulate methane monooxygenase gene pmoA and its use as a functional gene probe for methanotrophs. FEMS Microbiol Lett 156:205-210.

McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C*, et al.* (2006). The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci USA 103:15582-15587.

Meckenstock RU, Safinowski M, Griebler C (2004). Anaerobic degradation of polycyclic aromatic hydrocarbons. FEMS Microbiol Ecol 49:27-36.

Metzker ML (2009). Sequencing technologies—the next generation. Nature Reviews Genetics 11:31-46.

Mikkonen A, Kondo E, Lappi K, Wallenius K, Lindström K, Hartikainen H*, et al.* (2011). Contaminant and plant-derived changes in soil chemical and microbiological indicators during fuel oil rhizoremediation with *Galega orientalis*. Geoderma 160:336-346.

Moore FP, Barac T, Borremans B, Oeyen L, Vangronsveld J, Van der Lelie D*, et al.* (2006). Endophytic bacterial diversity in poplar trees growing on a BTEX-contaminated site: the characterisation of isolates with potential to enhance phytoremediation. Syst Appl Microbiol 29:539-556.

Mullins OC, Sheu EY, Hammami A, Marshall AG (2007). Asphaltenes, Heavy Oils, and Petroleomics. Springer: New York.

Mummey DL, Stahl PD (2003). Spatial and temporal variability of bacterial 16S rDNA-based T-RFLP patterns derived from soil of two Wyoming grassland ecosystems. FEMS Microbiol Ecol 46:113-120.

Muyzer G, Smalla K (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. A Van Leeuw 73:127-141.

Nedunuri K, Govindaraju R, Banks M, Schwab A, Chen Z (2000). Evaluation of phytoremediation for field-scale degradation of total petroleum hydrocarbons. J Environ Eng 126:483-490.

Newman LA, Strand SE, Choe N, Duffy J, Ekuan G, Ruszaj M*, et al.* (1997). Uptake and biotransformation of trichloroethylene by hybrid poplars. Environ Sci Technol 31:1062-1067.

Nocker A, Burr M, Camper AK (2007). Genotypic microbial community profiling: a critical technical review. Microb Ecol 54:276-289.

Oldenhuis R, Vink RL, Janssen DB, Witholt B (1989). Degradation of chlorinated aliphatic hydrocarbons by Methylosinus trichosporium OB3b expressing soluble methane monooxygenase. Appl Environ Microbiol 55:2819-2826.

Oliver MA, Webster R (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System 4:313-332.

Osborn AM, Moore ERB, Timmis KN (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ Microbiol 2:39-50.

Pace NR (1997). A molecular view of microbial diversity and the biosphere. Science 276:734-740.

Paisse S, Duran R, Coulon F, Goñi-Urriza M (2011). Are alkane hydroxylase genes (alkB) relevant to assess petroleum bioremediation processes in chronically polluted coastal sediments? Appl Microbiol Biotechnol 92:835-844.

Paliwal V, Raju SC, Modak A, Phale PS, Purohit HJ (2014). Pseudomonas putida CSV86: A Candidate Genome for Genetic Bioaugmentation. PLOS ONE 9:e84000.

Pérez-de-Mora A, Engel M, Schloter M (2011). Abundance and diversity of n-alkane-degrading bacteria in a forest soil co-contaminated with hydrocarbons and metals: a molecular study on alkB homologous genes. Microb Ecol 62:959-972.

Pérez-Pantoja D, Donoso R, Agulló L, Córdova M, Seeger M, Pieper DH*, et al.* (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. Environ Microbiol 14:1091-1117.

Petrov AA (1987). Petroleum hydrocarbons Springer-Verlag, New York, USA (1987), pp. 93–103.

Philippot L, Čuhel J, Saby N, Chèneby D, Chroňáková A, Bru D*, et al.* (2009). Mapping field-scale spatial patterns of size and activity of the denitrifier community. Environ Microbiol 11:1518-1526.

Pinyakong O, Habe H, Omori T (2003). The unique aromatic catabolic genes in sphingomonads degrading polycyclic aromatic hydrocarbons (PAHs). J Gen Appl Microbiol 49:1-19.

Podell S, Emerson JB, Jones CM, Ugalde JA, Welch S, Heidelberg KB*, et al.* (2014). Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. ISME J 8, 979–990.

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol 11:1358-1375.

Quince C, Curtis TP, Sloan WT (2008). The rational exploration of microbial diversity. ISME J 2:997-1006.

Ramos J, Duque E, Van Dillewjin P, Daniels C, Krell T, Espinosa-Urgel M*, et al.* (2010). Removal of hydrocarbons and other related chemicals via the rhizosphere of plants. In: Handbook of Hydrocarbon and Lipid Microbiology. Springer, pp 2575-2581.

Read D, Bengough A, Gregory P, Crawford JW, Robinson D, Scrimgeour C*, et al.* (2003). Plant roots release phospholipid surfactants that modify the physical and chemical properties of soil. New Phytol 157:315-326.

Rehm H, Reiff I (1981). Mechanisms and Occurrence of Microbial Oxidation of Long-Chain Alkanes. Adv Biochem Eng 19:175–216.

Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD*, et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283-290.

Rojo F (2009). Degradation of alkanes by bacteria. Environ Microbiol 11:2477-2490.

Roling WFM, Milner MG, Jones DM, Fratepietro F, Swannell RPJ, Daniel F*, et al.* (2004). Bacterial community dynamics and hydrocarbon degradation during a field-scale evaluation of bioremediation on a mudflat beach contaminated with buried oil. Appl Environ Microbiol 70:2603.

Röling WFM, Milner MG, Jones DM, Lee K, Daniel F, Swannell RJP*, et al.* (2002). Robust hydrocarbon degradation and dynamics of bacterial communities during nutrient-enhanced oil spill bioremediation. Appl Environ Microbiol 68:5537-5548.

Röling WFM, Ortega-Lucach S, Larter SR, Head IM (2006). Acidophilic microbial communities associated with a natural, biodegraded hydrocarbon seepage. J Appl Microbiol 101:290-299.

Ronaghi M, Uhlén M, Nyrén P (1998). A sequencing method based on real-time pyrophosphate. Science 281:363-365.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 242:84-89.

Rosenzweig AC (2011). Particulate methane monooxygenase. In: Handbook of Metalloproteins, Vols. 4 and 5 ed Messerschmidt A., editor. Chichester, UK: John Wiley & Sons, 615–622

Rossi RE, Mulla DJ, Journel AG, Franz EH (1992). Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecol Monogr 62:277-314.

Rothberg JM, Leamon JH (2008). The development and impact of 454 sequencing. Nat Biotechnol 26:1117-1124.

Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG*, et al.* (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4:1340-1351.

Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532-2538.

Schleheck D, Weiss M, Pitluck S, Bruce D, Land M, Han S*, et al.* (2011). Complete genome sequence of Parvibaculum lavamentivorans type strain (DS-1 T). Standards in genomic sciences 5.

Schleheck D, Tindall BJ, Rossello-Mora R, Cook AM (2004). *Parvibaculum lavamentivorans* gen. nov., sp. nov., a novel heterotroph that initiates catabolism of linear alkylbenzenesulfonate. Int J Syst Evol Microbiol 54:1489-1497.

Schloter M (2014). Influence of compost amendments on the diversity of alkane degrading bacteria in hydrocarbon contaminated soils. Front Microbiol 5:96.

Schnoor JL, Licht LA, McCutcheon SC, Wolfe NL, Carreira LH (1995). Phytoremediation of organic and nutrient contaminants. Environ Sci Technol 29:318A-323A.

Schnurer J, Rosswall T (1982). Fluorescein diacetate hydrolysis as a measure of total microbial activity in soil and litter. Appl Environ Microbiol 43:1256-1261.

Schulz S, Giebler J, Chatzinotas A, Wick LY, Fetzer I, Welzl G*, et al.* (2012). Plant litter and soil type drive abundance, activity and community structure of alkB harbouring microbes in different soil compartments. ISME J 6:1763-1774.

Schütte UM, Abdo Z, Bent SJ, Shyu C, Williams CJ, Pierson JD*, et al.* (2008). Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. Appl Microbiol Biotechnol 80:365-380.

Schwieger F, Tebbe CC (1998). A new approach to utilize PCR-single-strand-conformation polymorphism for 16S rRNA gene-based microbial community analysis. Appl Environ Microbiol 64:4870-4876.

Semrau JD, Chistoserdov A, Lebron J, Costello A, Davagnino J, Kenna E*, et al.* (1995). Particulate methane monooxygenase genes in methanotrophs. J Bacteriol 177:3071-3079.

Shakya M, Gottel N, Castro H, Yang ZK, Gunter L, Labbé J*, et al.* (2013). A Multifactor Analysis of Fungal and Bacterial Community Structure in the Root Microbiome of Mature Populus deltoides Trees. PloS one 8:e76382.

Shendure J, Ji H (2008). Next-generation DNA sequencing. Nat Biotechnol 26:1135-1145.

Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011). Pseudomonas genomes: diverse and adaptable. FEMS Microbiol Rev 35:652-680.

Singleton DR, Richardson SD, Aitken MD (2011). Pyrosequence analysis of bacterial communities in aerobic bioreactors treating polycyclic aromatic hydrocarbon-contaminated soil. Biodegradation 22:1061-1073.

Sipilä TP, Keskinen AK, Åkerman ML, Fortelius C, Haahtela K, Yrjälä K (2008). High aromatic ring-cleavage diversity in birch rhizosphere: PAH treatment-specific changes of IE 3 group extradiol dioxygenases and 16S rRNA bacterial communities in soil. ISME J 2:968-981.

Sipilä TP, Riisiö H, Yrjälä K (2006). Novel upper meta-pathway extradiol dioxygenase gene diversity in polluted soil. FEMS Microbiol Ecol 58:134-144.

Smalla K, Wieland G, Buchner A, Zock A, Parzy J, Kaiser S*, et al.* (2001). Bulk and Rhizosphere Soil Bacterial Communities Studied by Denaturing Gradient Gel Electrophoresis: Plant-Dependent Enrichment and Seasonal Shifts Revealed. Appl Environ Microbio 67:4742-4751.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR*, et al.* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". PNAS 103:12115-12120.

Söhngen N (1913). Benzin, petroleum, paraffinöl und paraffin als kohlenstoff-und energiequelle für mikroben. Zentralbl.Bacteriol.Parasitenk 37:595-609.

Song H, Bartha R (1990). Effects of jet fuel spills on the microbial community of soil. Appl Environ Microbiol 56:646-651.

Staley JT, Konopka A (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Reviews in Microbiology 39:321-346.

Stein ML (1999). Interpolation of Spatial Data: Some Theory for Kriging. New York: Springer

Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM*, et al.* (2004). A Populus EST resource for plant functional genomics. Proc Natl Acad Sci USA 101:13951-13956.

Stolz A (2013). Degradative plasmids from sphingomonads. FEMS Microbiol Lett 350: 9-19.

Story S, Parker S, Hayasaka S, Riley M, Kline E (2001). Convergent and divergent points in catabolic pathways involved in utilization of fluoranthene, naphthalene, anthracene, and phenanthrene by Sphingomonas paucimobilis var. EPA505. Journal of Industrial Microbiology and Biotechnology 26:369-382.

Susarla S, Medina VF, McCutcheon SC (2002). Phytoremediation: An ecological solution to organic chemical contamination. Ecol Eng 18:647-658.

Sutherland R, Van Kessel C, Pennock D (1991). Spatial variability of nitrogen-15 natural abundance. Soil Sci Soc Am J 55:1339-1347.

Suzuki M, Rappe MS, Giovannoni SJ (1998). Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl Environ Microbiol 64:4522-4529.

Taghavi S, Barac T, Greenberg B, Borremans B, Vangronsveld J, van der Lelie D (2005). Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene. Appl Environ Microbiol 71:8500-8505.

Tanaka S, Brentner LB, Merchie KM, Schnoor JL, Yoon JM, Aken BV (2007). Analysis of gene expression in poplar trees (Populus deltoides× nigra, DN34) exposed to the toxic explosive hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine (RDX). Int J Phytoremediation 9:15-30.

Taylor G (2002). Populus: Arabidopsis for forestry. Do we need a model tree? Annals of Botany 90:681.

Tervahauta AI, Fortelius C, Tuomainen M, Åkerman ML, Rantalainen K, Sipilä T*, et al.* (2009). Effect of birch (< i> Betula spp.) and associated rhizoidal bacteria on the degradation of soil polyaromatic hydrocarbons, PAH-induced changes in birch proteome and bacterial community. Environ Poll 157:341-346.

Tesar M, Reichenauer TG, Sessitsch A (2002). Bacterial rhizosphere populations of black poplar and herbal plants to be used for phytoremediation of diesel fuel. Soil Biol Biochem 34:1883-1892.

Thompson CC, Thompson FL, Vandemeulebroecke K, Hoste B, Dawyndt P, Swings J (2004). Use of recA as an alternative phylogenetic marker in the family Vibrionaceae. Int J Syst Evol Microbiol 54:919-924.

Throne-Holst M, Wentzel A, Ellingsen TE, Kotlar HK, Zotchev SB (2007). Identification of novel genes involved in long-chain n-alkane degradation by Acinetobacter sp. strain DSM 17874. Appl Environ Microbiol 73:3327-3332.

Törneman N, Yang X, Bååth E, Bengtsson G (2008). Spatial covariation of microbial community composition and polycyclic aromatic hydrocarbon concentration in a creosote-polluted soil. Environ Toxicol Chem 27:1039-1046.

Torsvik V, Ovreas L (2002). Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5:240-245.

Tringe SG, Hugenholtz P (2008). A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol 11:442-446.

Tuomi PM, Salminen JM, Jørgensen KS (2004). The abundance of nahAc genes correlates with the 14C-naphthalene mineralization potential in petroleum hydrocarbon-contaminated oxic soil layers. FEMS Microbiol Ecol 51:99-107.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U*, et al.* (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596-1604.

Uhlik O, Musilova L, Ridl J, Hroudova M, Vlcek C, Koubek J*, et al.* (2012). Plant secondary metabolite-induced shifts in bacterial community structure and degradative ability in contaminated soil. Appl Microbiol Biotechnol: 1-12.

Ulrich W, J Gotelli N (2007). Disentangling community patterns of nestedness and species co-occurrence. Oikos 116:2053-2061.

Vaillancourt FH, Bolin JT, Eltis LD (2006). The ins and outs of ring-cleaving dioxygenases. Crit Rev Biochem Mol Biol 41:241-267.

Van Aken B, Yoon JM, Just CL, Schnoor JL (2004). Metabolism and mineralization of hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine inside poplar tissues (Populus deltoides× nigra DN-34). Environ Sci Technol 38:4572-4579.

Van Beilen JB, Li Z, Duetz WA, Smits THM, Witholt B (2003). Diversity of alkane hydroxylase systems in the environment. Oil & gas science and technology 58:427-440.

Van Beilen JB, Funhoff EG (2007). Alkane hydroxylases involved in microbial alkane degradation. Appl Microbiol Biotechnol 74:13-21.

Van Beilen JB, Marin MM, Smits TH, Röthlisberger M, Franchini AG, Witholt B*, et al.* (2004). Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium Alcanivorax borkumensis. Environ Microbiol 6:264-273.

Van Beilen JB, Li Z, Duetz WA, Smits TH, Witholt B (2003). Diversity of alkane hydroxylase systems in the environment. Oil & gas science and technology 58:427-440.

Van Beilen JB, Funhoff EG, van Loon A, Just A, Kaysser L, Bouza M*, et al.* (2006). Cytochrome P450 alkane hydroxylases of the CYP153 family are common in alkane-degrading eubacteria lacking integral membrane alkane hydroxylases. Appl Environ Microbiol 72:59-65

Van Beilen JB, Panke S, Lucchini S, Franchini AG, Rothlisberger M, Witholt B (2001). Analysis of Pseudomonas putida alkane-degradation gene clusters and flanking insertion sequences: evolution and regulation of the alk genes. Microbiology 147:1621-1630.

Van Dillewijn P, Couselo JL, Corredoira E, Delgado A, Wittich R, Ballester A*, et al.* (2008). Bioremediation of 2, 4, 6-trinitrotoluene by bacterial nitroreductase expressing transgenic aspen. Environ Sci Technol 42:7405-7410.

Van Elsas J, Boersma F (2011). A review of molecular methods to study the microbiota of soil and the mycosphere. Eur J Soil Biol 47:77-87.

Vervaeke P, Luyssaert S, Mertens J, Meers E, Tack F, Lust N (2003). Phytoremediation prospects of willow stands on contaminated sediment: a field trial. Environ Poll 126:275-282.

Vidali M (2001). Bioremediation. An overview. Pure Appl Chem 73:1163-1172.

Vieira S (1983). Geostatistical theory and application to variability of some agronomical properties. California.University, Berkeley.Division of Agricultural Sciences.Hilgardia 51.

Vitte I, Duran R, Hernandez-Raquet G, Mounier J, Jézéquel R, Bellet V*, et al.* (2013). Dynamics of metabolically active bacterial communities involved in PAH and toxicity elimination from oil-contaminated sludge during anoxic/oxic oscillations. Appl Microbiol Biotechnol :1-13.

Vorobev AV, Baani M, Doronina NV, Brady AL, Liesack W, Dunfield PF*, et al.* (2011). Methyloferula stellata gen. nov., sp. nov., an acidophilic, obligately methanotrophic bacterium that possesses only a soluble methane monooxygenase. Int J Syst Evol Microbiol 61:2456-2463.

Walker LR, Del Moral R (2003). Primary Succession and Ecosystem Rehabilitation. Cambridge University Press.

Walker TS, Bais HP, Grotewold E, Vivanco JM (2003). Root exudation and rhizosphere biology. Plant Physiol 132:44-51.

Wang W, Shao Z (2012). Diversity of flavin-binding monooxygenase genes (*almA*) in marine bacteria capable of degradation long-chain alkanes. FEMS Microbiol Ecol 80:523-533.

Wang X, Bartha R (1990). Effects of bioremediation on residues, activity and toxicity in soil contaminated by fuel spills. Soil Biol Biochem 22:501-505.

Wang LT, Lee FL, Tai CJ, Kasai H (2007). Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group. Int J Syst Evol Microbiol 57:1846-1850.

Wentzel A, Ellingsen TE, Kotlar H, Zotchev SB, Throne-Holst M (2007). Bacterial metabolism of long-chain n-alkanes. Appl Microbiol Biotechnol 76:1209-1221.

Wertz S, Degrange V, Prosser JI, Poly F, Commeaux C, Guillaumaud N*, et al.* (2007). Decline of soil microbial diversity does not influence the resistance and resilience of key soil microbial functional groups following a model disturbance. Environ Microbiol 9:2211-2219.

Wessén E, Söderström M, Stenberg M, Bru D, Hellman M, Welsh A*, et al.* (2011a). Spatial distribution of ammonia-oxidizing bacteria and archaea across a 44-hectare farm related to ecosystem functioning. ISME J 5:1213-1225.

Weyens N, Van Der Lelie D, Artois T, Smeets K, Taghavi S, Newman L*, et al.* (2009). Bioaugmentation with engineered endophytic bacteria improves contaminant fate in phytoremediation. Environ Sci Technol 43:9413-9418.

White JC, Newman LA (2011). Phytoremediation of Soils Contaminated with Organic Pollutants. Biophysico-Chemical Processes of Anthropogenic Organic Compounds in Environmental Systems :503-516.

Whyte L, Schultz A, Beilen J, Luz A, Pellizari V, Labbé D*, et al.* (2002). Prevalence of alkane monooxygenase genes in Arctic and Antarctic hydrocarbon-contaminated and pristine soils1. FEMS Microbiol Ecol 41:141-150.

Will C, Thürmer A, Wollherr A, Nacke H, Herold N, Schrumpf M*, et al.* (2010). Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol 76:6751-6759.

Woese CR (1987). Bacterial evolution. Microbiol Rev 51:221.

Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K*, et al.* (1975). Conservation of primary structure in 16S ribosomal RNA.

Woese CR, Kandler O, Wheelis ML (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576-4579.

Woese CR, Fox GE (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088-5090.

Wollum A, Cassel D (1984). Spatial variability of Rhizobium japonicum in two North Carolina soils. Soil Sci Soc Am J 48:1082-1086.

Wullschleger SD, Jansson S, Taylor G (2002). Genomics and forest biology: Populus emerges as the perennial favorite. Plant Cell 14:2651-2655.

Yadav JS, Loper JC (1999). Multiple P450alk (cytochrome P450 alkane hydroxylase) genes from the halotolerant yeast< i> Debaryomyces hansenii. Gene 226:139-146.

Yateem A, Al-Sharrah T, Bin-Haji A (2007). Investigation of microbes in the rhizosphere of selected grasses for rhizoremediation of hydrocarbon-contaminated soils. Soil and Sediment Contamination: An International Journal 16:269-280.

Yergeau E, Sanschagrin S, Maynard C, St-Arnaud M, Greer CW (2014). Microbial expression profiles in the rhizosphere of willows depend on soil contamination. ISME J 8:344-358.

Yrjälä K, Keskinen AK, Åkerman ML, Fortelius C, Sipilä TP (2010). The rhizosphere and PAH amendment mediate impacts on functional and structural bacterial diversity in sandy peat soil. Environmental Pollution 158:1680-1688.

Yrjälä K, Tuomivirta T, Juottonen H, Putkinen A, Lappi K, Tuitilla E*, et al.* (2011). CH4 production and oxidation processes in a boreal fen ecosystem after long-term water table drawdown. Global Change Biol 17:1311-1320.

Yu S, Li S, Tang Y, Wu X (2011). Succession of bacterial community along with the removal of heavy crude oil pollutants by multiple biostimulation treatments in the Yellow River Delta, China. Journal of Environmental Sciences 23:1533-1543.

Zeng Y, Feng F, Liu Y, Li Y, Koblížek M (2013). Genome Sequences and Photosynthesis Gene Cluster Composition of a Freshwater Aerobic Anoxygenic Phototroph, Sandarakinorhabdus sp. Strain AAP62, Isolated from the Shahu Lake in Ningxia, China. Genome Announcements 1.

Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD*, et al.* (2014). Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proc Natl Acad Sci USA 111:E836-45.

Zuckerkandl E, Pauling L (1965). Molecules as documents of evolutionary history. J Theor Biol 8:357-366.